REMARKS

Claims 27-36 are pending in the present application.

The Specification is amended to correctly specify the priority data of the instant application under the section "Cross-Reference to Related Applications." No new matter is added and entry of the amendments to the Specification and claims is respectfully requested.

Reconsideration of the application is respectfully requested in view of the above amendments and the following remarks. For the Examiner's convenience, Applicant's remarks are presented in the order in which they were raised in the Office Action.

A. Non-statutory Double Patenting Rejection

(a) Claims 27-30 stand rejected under the judicially created doctrine of obviousness-type double patenting as being unpatentable over claims 1-4 of U.S. Patent No. 5,585,258.

Claims 31-35 are rejected under the judicially created doctrine of obviousness-type double patenting as being unpatentable over claims 5-9 of U.S. Patent No. 5,585,258 in view of Benson et al., U.S. Patent No. 5,258,496. Benson is cited for the teaching of recombinant fusion polyupeptides being comprised in compositions during purification from the host cell.

Claim 36 is rejected under the judicially created doctrine of obviousness-type double patenting as being unpatentable over claims 1 and 3-5 of U.S. Patent No. 5,597,691.

Claims 27 and 30 are rejected under the judicially created doctrine of obviousness-type double patenting as being unpatentable over claims 1 and 2 of U.S. Patent No. 5,712,145.

Claims 31, 32 and 35 are rejected under the judicially created doctrine of obviousness-type double patenting as being unpatentable over claims 3-5 of U.S. Patent No. 5,712,145 in view of Benson et al., U.S. Patent No. 5,258,496.

Claim 36 is rejected under the judicially created doctrine of obviousness-type double patenting as being unpatentable over claims 7 and 8 of U.S. Patent No. 5,712,145.

Applicants submit that they will file a terminal disclaimer in the present application to disclaim any term beyond the term of the earlier expiring patents in order to overcome this ground for rejection, after the conflicting claims are found to be allowable.

(b) Claims 27 and 30 are provisionally rejected under the judicially created doctrine of obviousness-type double patenting as being unpatentable over claim 11 of copending Application No. 10/409,094, which is an application for reissue of U.S. Patent No. 5,585,258.

Applicants submit that they will file a terminal disclaimer in the appropriate case -- the present application or copending Application No. 10/409,094 – to disclaim any term beyond the term of the earlier expiring patent in order to overcome this ground for rejection, after the conflicting claims are found allowable.

Claim 36 is provisionally rejected under the judicially created doctrine of obviousness-type double patenting as being unpatentable over claim 6 of copending Application No. 10/409,673, which is an application for reissue of U.S. Patent No. 5,597,691.

Applicants submit that they will file a terminal disclaimer in the appropriate case -- the present application or copending Application No. 10/409,673 – to disclaim any term beyond the term of the earlier expiring patent in order to overcome this ground for rejection, after the conflicting claims are found allowable.

B. Rejections under 35 USC § 112

1. Rejections under 35 U.S.C. §112, first paragraph – written description

Claims 27, 31, and 36 (and dependent claims thereof) stand rejected under 35 U.S.C. § 112, first paragraph for lack of written description, as containing subject which was not described in the Specification in such a way as to reasonably convey to one skilled in the relevant art that the inventors, at the time the application was filed, had possession of the claimed invention.

Specifically, the Examiner contends that the Specification fails to exemplify or describe the preparation of, or recite any structural features of, the "proteolytic hepatitis C virus polypeptides" recited by the claims. (Office Action at p. 5). The Examiner states that the Specification fails to identify an amino acid sequence that constitutes a proteolytic HCV polypeptide or describe its purification. *Id.* The Examiner also states that the application fails to disclose a domain protease with proteolytic activity and contends that Example 5 of the Specification suggests that the proteolytic products detected by ELISA could only have been produced by endogenous proteases. The Examiner further states that no "purified proteolytic

HCV peptide" of claim 36 is shown to cleave any peptide substrate and no identifying characteristics of a generic NS3 domain protease is shown.

Applicants respectfully traverse for the following reasons:

(i) A Written Description of a "proteolytic hepatitis C virus polypeptide" that comprises a hepatitis C virus NS3 domain protease or an active NS3 domain hepatitis C virus protease truncation analog is provided.

"[T]he 'essential goal' of the description of the invention requirement is to clearly convey the information that an applicant has invented the subject matter which is claimed." *In re Barker*, 559 F.2d 588, 592 n.4, 194 USPQ 470, 473 n.4 (CCPA 1977). The test for sufficiency of support in a parent application is whether the disclosure of the application relied upon "reasonably conveys to the artisan that the inventor had possession at that time of the later claimed subject matter." *Ralston Purina Co. v. Far-Mar-Co., Inc.*, 772 F.2d 1570, 1575, 227 USPQ 177, 179 (Fed. Cir. 1985) (quoting *In re Kaslow*, 707 F.2d 1366, 1375, 217 USPQ 1089, 1096 (Fed. Cir. 1983)).

(a) An NS3 domain is described in the Specification

The Specification states that: "[t]he term 'HCV protease' refers to an enzyme derived from HCV which exhibits proteolytic activity, specifically the polypeptide encoded in the NS3 domain of the HCV genome." (Specification, page 6, lines 22-25) An HCV NS3 domain protease sequence is provided in Figure 1 of the Specification. (Specification, page 3, line 7). The Specification points to a specific section in the NS3 domain as the key to proteolytic activity and notes that the termini of the relevant section are putative. (Specification, page 6, line 24 through page 7, line 21). The Specification describes an NS3 domain of HCV. Page 8, lines 7-25 refer to NS3 domain by analogy with the Yellow Fever Virus (a flavivirus) polyprotein. An HCV protease encoded by the NS3 domain in at least one strain of HCV is further described with reference to a 202 amino acid protease sequence from SEQ ID NO: 1 in page 6, line 22 to page 7, line 18 (*see* SEQ ID NO: 65).

An "active" truncation analog is one that exhibits proteolytic activity, a property that one can ascertain by running a limited number of standard experiments. The Specification describes how one would determine the structure of the shortest active HCV NS3 protease by truncation analysis. (Specification, page 7, line 27 – page 8, line 6).

In a European patent application EP 318,216A1 (published May 31, 1989), the inventors of the present application had previously reported[1] the nucleotide sequence of the HCV genome and identified a similarity between a 530 amino acid domain of the HCV polyprotein sequence and the NS3 protein sequence of dengue virus, a flavivirus. (p. 52, sec. IV.H.3 and Figs. 41-1 and 41-2). Likewise, in PCT application WO 89/04669 a correlation between HCV polyprotein and a nonstructural protein of the flavivirus was noted. (p. 128, sec. IV.H.3). The disclosures of both WO 89/04669 and EP 318,216 (Houghton et al.) are incorporated by reference in the instant Specification at page 4, lines 4-8.

NS3 domains in flaviviruses such as yellow fever virus were known in the art. (*see* Fig. 1 and page 731 in Rice CM *et al.*, Science, 229(4715):726-733 (1985)). EP 388,232 by the same inventive entity as the current application and published September 19, 1990, identified the NS3 domain in comparison with flaviviruses. (pages 33-34 of EP 388,232). Other publications identifying an NS3 domain protease of HCV were available prior to the filing of the priority application of the current application. Computer aided comparative analysis of the polyproteins of several flaviviruses was known to have sequence similarity with HCV in the NS3 region. (Miller *at al.* Proc. Natl. Acad. Sci. 87:2057-2061, at 2060 and Fig. 3 (March 1990)). Yoneyama *et al.* disclose the use of PCR primer from the NS3 region of HCV for detection of viral sequences. (Jpn. J. Med. Sci. Biol. 43:89-94 (1990)).[2]

(b) An NS3 domain protease is disclosed in the Specification.

Independent claims 27 and 36 specify a composition comprising a purified hepatitis C virus polypeptide which itself comprises "an HCV NS3 domain protease or an active HCV NS3 domain protease truncation analog." Independent claim 31 specifies a composition comprising a purified hepatitis C virus polypeptide which itself comprises a fusion protein containing "a HCV NS3 domain protease or . . . an active HCV NS3 domain protease truncation analog." The remaining dependent claims are generally limited to truncation analogs containing the amino acid sequence of SEQ ID NOS: 63-65. The claims are not directed to a specific kind of protease activity, they are directed to *any* protease activity encoded by the NS3 region.

---

[1] EP 318,216A1 was filed on Nov. 11, 1988 and published May 31, 1989, prior to the earliest priority date of this application.

[2] Courtesy copies of the references mentioned in the response are attached as Exhibit H for the Examiner's convenience. Only the relevant pages of EP 388,232, WO 89/04669 and EP 318,216 are enclosed.

A protease activity associated with the NS3 domain is characterized in Example 5 (Specification, page 31, lines 12-17) which shows self cleavage of SOD - HCV protease fusion proteins expressed in *E. coli*.[3] Example 4 (Specification, page 29, line 4 through page 30, line 6) describes the amino acids of HCV protease encoded by each fusion protein.

- The P190 fusion product encoding amino acids 1-199 of the HCV protease (page 29, lines 19-20) showed no protease cleavage activity (Specification, page 32, lines 8-12).

- P300 which includes amino acids 1-299 of HCV protease (page 29, lines 25-26) indicated occurrence of cleavage (Specification, page 32, lines 1-7).

- P500 comprising amino acids 1-513 of Fig. 1 (page 30, lines 4-6) indicated occurrence of cleavage (Specification, page 31, lines 22-25).

- The fusion protein ("P600") encoded by the vector cf1SODp600 which includes amino acids 1-686 of Fig. 1 also showed proteolytic activity. (Specification, page 31, lines 12-17).

- The Specification concludes that "the minimum essential sequence for HCV protease extends to the region between amino acids 199 and 299." (Specification, page 32, lines 10-12).

The Examiner incorrectly assumes that the proteolytic cleavage described in Example 5 is attributable *solely* to the host cells' endogenous proteases. Only in subsection A of Example 5 which describes the protease activity of the P600 fusion protein resulting in "34, 53 and 66 kDa" bands, the 53 and 66 kDa bands are surmised to have undergone "varying degrees of (possibly bacterial) processing" as the predicted product of theoretical M_r = 93 kDa was not observed. (Specification, page 31, lines 13-17).

Protease activity attributable to the NS3 region was evident in the P300 and P500 fusion proteins and no "possibly bacterial" processing is suggested in Examples 5 (B) and (C) as the predicted proteolysis products of theoretical M_r = 51 and 73 kDa respectively were observed. (Specification, page 31, lines 22-25, and page 32, lines 1-2). That the protease activity resides in

---

[3] "The results indicated the occurrence of cleavage, as no full length product (theoretical M_r = 93 kDa) was evident on the gel." (Specification, page 31, lines 12-13).

the HCV NS3 region is confirmed by the observation in Example 5(C), where the P190 fusion product encoding amino acids 1-199 of the HCV protease did not show any protease cleavage activity (Specification, page 32, lines 8-12).

(c) A peptide substrate for the NS3 domain protease is provided in the Specification

The Examiner further contends that the application fails to disclose a proteolytic HCV peptide encoded by a polynucleotide or expression vector that is capable of encoding that could cleave any peptide substrate. Applicants respectfully traverse and submit that a peptide substrate for a HCV NS3 associated protease is disclosed in the Specification. The protease activity described in Examples 5(A), (B), and (C) was observed through self-cleavage of an hSOD-HCV fusion protein wherein the HCV peptide portion corresponded to amino acids 1-686 of Fig. 1 and various truncations thereof. Observance of specific cleavage within the NS3 region is described in every instance where protease activity was observed. For example, "34 kDa band correspond[ing] to the hSOD partner (about 20 kDa) with a portion of the NS3 domain" was observed in each case with the P600, P300 and P500 fusion proteins of NS3 fused to a hSOD leader.

Applicants submit that the Specification describes a protease activity specifically associated with the NS3 region and provides disclosure of a substrate for such protease activity. Thus one of skill in the art would have identified the NS3 domain described in the Specification and understood that at the time of filing of the application, the inventors had possession of the claimed invention. Therefore, applicant respectfully requests withdrawal of this ground for rejection for lack of written description under 35 U.S.C. § 112, first paragraph.

(d) NS4A is not essential for the activity of an NS3 domain hepatitis C virus protease or truncation analog

The Examiner refers to several references submitted in an IDS by the Applicants to contend that NS3 domain hepatitis C virus protease requires another region termed NS4A. Applicants respectfully traverse.

The claims of the current application generally specify a composition comprising a purified hepatitis C virus polypeptide which itself comprises "an HCV NS3 domain protease or an active HCV NS3 domain protease truncation analog." The claims are not directed to a specific kind of protease activity but any protease activity encoded by the NS3 region. The

claims do not specify a particular kind of protease. The NS4A cofactor referred by the Examiner relates to the activity of a "serine protease activity" encoded by the NS3 region. The Specification clearly demonstrates a protease activity associated with a protein comprising amino acids 1-299 of HCV protease (*see* Example 5 (C)). While a serine protease activity also encoded within this region may optionally require a NS4A cofactor, applicants' claims are directed to any protease activity within the NS3 region. Further, applicants note that "while NS4A appears to be absolutely required for *trans*-cleavage at the 4B/5A site, it is not an essential cofactor for serine protease activity." (*see* Abstract, lines 10-11, page 8151 right column (first full paragraph) of Lin *et al.*, J. Virol. 68(12): 8147-8157 (1994)). Further *cis*-cleavage by NS3 domain proteases do not require NS4a. (Lin *et al.* p. 8149, right col.; p. 8152, right col.; Fig. 7A; p. 8155, left col.).

Applicants submit that the Specification describes a protease activity specifically associated with the NS3 region and provides disclosure of a substrate for such protease activity. Thus one of skill in the art would have identified the NS3 domain described in the Specification and understood that at the time of filing of the application, the inventors had possession of the claimed invention. Therefore, applicants respectfully request withdrawal of this ground for rejection for lack of written description under 35 U.S.C. § 112, first paragraph.

2. Rejections under 35 U.S.C. §112, First paragraph – enablement

Claims 27-36 are rejected under 35 U.S.C. § 112, first paragraph, because the Specification, while being enabling for recombinant expression of a catalytic component of a hepatitis C virus protease comprising the amino acid sequence set forth in SEQ ID NO:66, does not reasonably provide enablement for preparation of compositions, or expression vectors, that comprise polynucleotides encoding a proteolytically active hepatitis C virus protease, whether or not fused to a fusion partner. The Specification does not enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and use the invention commensurate in scope with these claims.

Specifically, the Examiner contends that while appropriate analogies are made in the Specification between SEQ ID NO: 66, serine protease characteristics, and analogous regions in other flaviviruses, no guidance is provided for making such a protease. The Examiner states that the Specification "does not describe, thus cannot enable, an integral hepatitis C virus protease capable of cleaving a defined substrate." (Office Action, page 8). The Examiner also states that

the small peptides specified in claims 28, 29, 33 and 34 (*i.e.*, of SEQ ID NOS: 63 and 64) "are insufficient to support proteolysis even if Applicants' disclosure had provided [such] guidance. *Id.*

Applicants respectfully traverse these grounds for rejection.

(i) The Specification enables a NS3 domain protease of the amino acid sequences of SEQ ID NOS: 63, 64, 65 and 66

To be enabling, the Specification must teach those skilled in the art how to make and use the full scope of the claimed invention without undue experimentation. (*Genentech Inc. v. NovoNordisk A/S*, 108 F.3d 1361, 42 USPQ2d 1001 (Fed. Cir. 1997)).

(a) Applicants submit that the Wands factor cited by the Examiner, *i.e.*, specific guidance about the portion of the HCV polyprotein responsible for recognition of native cleavage sites in the polyprotein, is satisfied in the Specification. Specifically the truncation analysis described in Example 5, wherein a minimal region between 199-299 amino acids of Fig. 1 is shown to have protease activity with specific cleavage occurring within the HCV NS3 portion of the hSOD-HCV fusion protein.

As submitted above, the Specification identifies a HCV NS3 region. As described in Example 5, a protease activity is shown to be associated with amino acids 1-299 of Fig. 1 (P300 fusion protein; SEQ ID NO: 66). Example 5 also shows that no protease activity is observed within amino acids 1-199 of Fig. 1 (P190 fusion protein; SEQ ID NO:67). SEQ ID NO: 65 extends from amino acid 60-262 of Fig. 1. Thus the amino acid sequence essential for the protease activity is located within amino acids 200 and 262 of the given sequence. One of skill in the art has only a definite and specific region of the amino acid sequence to identify the protease activity and is able to do so without undue experimentation.

SEQ ID NOS: 63 and 64 specify 11 and 9 amino acid sequences within SEQ ID NO: 65 and are specified in dependent claims 28 and 29 which depend from independent claim 27, and claims 33 and 34 which depend from independent claim 31. Independent claims 27 and 31 both specify a "proteolytic hepatitis C virus (HCV) polypeptide wherein said HCV polypeptide comprises .... [a] protease." (emphasis added). The proteolytic HCV polypeptides according to claims 28, 29, 33 and 34 need only have "a partial internal amino acid sequence comprising" SEQ ID NOS: 63 and 64.

SEQ ID NOS: 63 and 64 span 11 and 9 amino acid sequences within SEQ ID NO:65 and are within the protease domain of amino acids 200 and 262 of the given sequence of Fig. 1. Further, SEQ ID NOS: 63 and 64 span a histidine and a serine containing region respectively of sequences homologous to regions responsible for serine protease catalytic activity in Yellow Fever Virus, West Nile Fever virus, Murray Valley Fever virus, and Kunjin virus (Table 1) and in the well-characterized serine proteases: protease A from *Streptomyces griseus*, α-lytic protease, bovine trypsin, chymotrypsin, and elastase (Table 2). (*see* page 8, line 7 – page 9, line 17). Thus, by structural homology and alignment, SEQ ID NOS: 63 and 64 are disclosed in the Specification to be associated with protease activity.

While the Specification notes characteristic similarities with a serine protease the claims are directed more broadly to a "protease" activity within the NS3 domain. Applicants are not required to correctly set forth, or even know, how and why the claimed NS3 region demonstrates protease activity. *see Enzo Biochem v. Calgene, Inc.*, 188 F.3d 1362, 1375 (Fed. Cir. 1999) ("it is not a requirement of patentability that an inventor correctly set forth, or even know, how or why the invention works").

(b) The Examiner also alleges a lack of working examples of an assay that could measure inhibition of an HCV protease. Applicants respectfully traverse.

As discussed above, the Specification discloses an NS3 domain with protease activity residing in a region defined by truncation analysis. Examples 4 and 5 discloses several hSOD-protease fusion proteins (cf1SODp600, P300, P500) that act specifically as substrates of the NS3 domain protease. Given a protease and a substrate one of skill in the art would be able to assay for inhibitors of the protease activity. Inhibitors such as organic compounds, peptide inhibitors and antibodies and methods for designing them are disclosed on pages 17-18 of the Specification. General methods for screening protease inhibitors are listed at pages 18-19 of the Specification.

(c) The Examiner also states that "in view of the publications made of record herein," the existing state of the art at the time of filing does not support the identification of other, distant regions of flavivirus sequence that confer cleavage specificity. Applicants respectfully traverse. As discussed above in Section 1(i)(d) the Specification describes a protease activity specifically associated with the NS3 region. The claims are not directed to a specific kind of protease activity but any protease activity encoded by the NS3 region. Applicants submit that the Specification

describes a protease activity specifically associated with the NS3 region and provides disclosure of a substrate for such protease activity. (*see* discussion under section 1(i)(d) above).

Applicants respectfully request that in view of the description of an NS3 domain, a protease activity associated with the NS3 domain, and identification of a substrate specifically cleaved by the NS3 protease in the Specification, the rejection for lack of enablement under 35 U.S.C. § 112, first paragraph be withdrawn.

3. Rejections under 35 U.S.C. §112, Second paragraph – indefiniteness

a. Indefiniteness of the terms "domain protease" and "truncation analog"

Claims 27-36 stand rejected under 35 U.S.C. § 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which applicant regards as the invention. Specifically, the Examiner concludes that independent claims 27 and 31 are indefinite in reciting, "proteolytic hepatits C virus polypeptide . . . compris[ing] an HCV NS3 domain protease or an active . . . truncation analog," because the Specification does not provide a specific, limiting, structural description of a generic NS3 domain protease, and thus one could not determine what is more than the protease and what is a truncation analog.

Applicants respectfully traverse. The application, at page 5, line 20 through page 6, line 4, refers to NS3 domain by analogy with the Yellow Fever Virus polyprotein. An HCV protease encoded by or within the NS3 domain is further described with reference to a 202 amino acid protease within SEQ ID NO: 1 in at least one strain of HCV in page 6, line 26 though page 7, line 18. SEQ ID NO: 65 consists of the corresponding 202 amino acid sequence from Figure 1 (amino acids 60-262).

The protease activity associated with HCV NS3 domain is further characterized in Example 5 (Specification, pages 31-32) as discussed in detail above. The Specification identifies by truncation analysis described in Example 5 a "minimum essential sequence for HCV protease [that] extends to the region between amino acids 199 and 299." (Specification, page 32, lines 10-12).

Further, Examples 4 and 5, as discussed above, disclose active truncation analogs of the HCV NS3 domain protease. The Specification on pages 29-32 disclose a fusion protein P600 including amino acids 1-686 of Fig. 1 which demonstrates protease activity. Active truncation

analogs P500 comprising amino acids 1-513 of Fig. 1, and P300 comprising amino acids 1-299 also demonstrate protease activity associated with the NS3 domain. (Specification, page 31, line 5 – page 32, line 7).[4]

Applicants submit that the terms "HCV NS3 domain protease" and "active ... truncation analog" are clearly defined in the Specification and request withdrawal of this ground for rejection.

b. Indefiniteness of the term "purified"

The Examiner also states that claims 27 and 31 (and dependent claims) are indefinite because they recite a composition comprising a purified polypeptide, where no polypeptide can remain "purified" when present in a composition. Claim 36 is likewise indefinite because it recites a method involving the step of providing a purified polypeptide. According to the Examiner, the claims also do not provide for purification of proteolytic HCV polypeptides.

Applicants respectfully traverse the Examiner's interpretation of a composition comprising a "purified" polypeptide. Claims 27 and 31 specify compositions comprising the "purified" HCV polypeptides as a component of the compositions. Absolute purity of HCV polypeptides within the compositions are not claimed.

The ordinary meaning of "purify" is to "free from undesirable elements." (Merriam-Webster's Collegiate Dictionary, 10th ed. 2002. Merriam-Webster, Inc. Springfield, MA; Exhibit E) Therefore, a "purified" substance can exist in a composition so long as it is free from undesirable elements. Accordingly, a "purified" proteolytic HCV polypeptide can exist in a composition.

Further, the Specification uses the term "purified" in a manner consistent with this meaning. For instance, it discloses that a calcium dependent monoclonal antibody, which binds to the FLAG encoded peptide, can be used to purify the fusion protein without harsh eluting conditions. (Page 31, lines 2-3). The "purified" protein is obtained as an antibody-protein complex, a composition comprising a purified protein. Similarly, Example 6 of the Specification describes a method of purifying protease from *E. coli* by SDS-PAGE. The final purified product

[4] Truncation analog P190 comprising amino acids 1-199 does not demonstrate proteolytic activity, and is thus an "inactive" truncation analog. (Specification, page 32, lines 8-10).

is eluted from a gel. (Page 33, lines 20-23). In this instance, the "purified" protease is part of a composition comprising the eluate.

Therefore, Applicants respectfully request that the rejection for indefiniteness under 35 U.S.C. § 112, first paragraph be withdrawn.

C. Rejections under 35 USC § 103

Claims 27-35 stand rejected under 35 U.S.C. § 103(a) as being unpatentable over Miyamura et al., U.S. 5,372,928, in view of Miller et al., 1990, Proceedings of the National Academy of Sciences, U.S.A, Vol. 87, pages 2057-2061; Bazan et al., 1989, Virology, Vol. 171, pages 637-639 and Gorbalenya et al., 1989, Nucleic Acids Research, Vol. 17, pages 3889-3897.

Applicants respectfully traverse this rejection because it does not establish *prima facie* obviousness of the claimed inventions. In particular, the key teachings in the Miyamura patent relied upon in the rejection are not available as prior art against the claimed inventions as a matter of law. Since the Miyamura patent is the primary reference, the rejection should be withdrawn.

The Office Action characterizes the Miyamura patent as follows:

(1) Miyamura et al., see Figures 12A-C, teach a polynucleotide encoding the HCV1 strain polyprotein and the relative positions of both the structural and the non-structural domains within the polyprotein encoded by the nucleic acid sequence of the HCV1 strain, see Miyamura patent, columns 6-7 and Figure 11. Office Action at 10.

(2) Miyamura et al. teach, at lines 8-10 of col. 7, that the "putative NS3 [domain extends] from about amino acid 1007 to about amino acid 1650'." *Id.* (text within brackets in original).

(3) Miyamura et al. teach, at col. 17, lines 5-21, that functions of domains within the HCV polyprotein may be predicted on the basis of similarities shared by the HCV polyprotein amino acid sequence and flavivirus polyprotein amino acid sequences and that flavivirus NS3 domains have an amino acid sequence region that provides a protease function. Office Action at 10-11. According to the Examiner, these teaching have priority to September 15, 1989, when they appeared in the parent application serial No. 07/408,045.

(4) At Examples I-IV at cols. 28-39, Miyamura et al. "teach preparation of cloning vectors, and transformed host cells comprising the vectors, comprising inserts of specific, defined, regions found anywhere in a nucleic acid sequence encoding all or part of an hepatitis C virus polyprotein." Office Action at 11. Miyamura et al. "explicitly teach, at cols. 8-10, that expression vectors comprising transcriptional and translational regulatory elements operably linked to a polynucleotide encoding a desired regions [sic] of the HCV polyprotein should be used to produce desired portions of the hepatitis C virus polyprotein in host cells." *Id.* Miyamura et al. "further suggest preparation of expression constructs providing fusions of hepatitis C virus amino acid sequence regions with proteins commonly used in the art as fusion partners such as β-galactosidase and superoxide dismutase [SOD]" at cols. 14-15. Office Action at 11 (bracketed text in original).

1. The Instant Application Claims Priority to U.S. Patent No. 5,371,017

The instant application has priority of Application No. 07/680,296, filed on April 4, 1991, (now U.S. Patent No. 5,371,017) and its Specification is identical to that of the '017 patent.[5] The inventors of the instant application and the '017 patent are identical.

As discussed below, Miyamura et al. is not available as § 102(e) prior art against the '017 patent. Since the instant application claims priority (and contains identical disclosure) to the application from which the '017 patent issued, Miyamura is not available as prior art against the instant application.

2. The Miyamura Patent Cannot be Relied Upon as § 102(e) Prior Art With Respect to the HCV-1 ORF Sequence Information Shown in Figure 12, the Putative Genomic Organization Shown in Figure 11, and the Subject Matter of Column 7, Lines 8-9 and Column 17, Lines 17-21 of the Miyamura Patent

Subject matter disclosed in the Miyamura patent qualifies as prior art under 35 U.S.C. § 103 only if it meets the requirements of 35 U.S.C. § 102(e). The Office Action relies in part on Figure 12, Figure 11 and "cols. 6-7" of the Miyamura patent, stating that the patent teaches a

---

[5] The instant application, 09/884,455, is a continuation of Application No. 09/253,675, which is a continuation of Application No. 08/709,177 (U.S. Patent No. 5,885,799), which is a continuation of Application No. 08/440,548 (U.S. Patent No. 5,597,691), which is a divisional of Application No. 08/350,884 (U.S. Patent No. 5,585,258), which is a divisional of Application No. 07/680,296, filed April 4, 1991, (U.S. Patent No. 5,371,017).

polynucleotide encoding the HCV1 strain's polyprotein and "the relative positions of the structural and the non-structural domains within the polyprotein." Office Action at 10. The Office Action also relies in part on the Miyamura patent at column 7, lines 8-10, asserting that it teaches there that the "putative NS3 [domain extends] from about amino acid 1007 to about amino acid 1650" (bracketed text in original). *Id.* The Office Action also relies in part on the Miyamura patent at column 17, lines 5-21, stating that the patent there teaches "that functions of domains within the hepatitis C virus polyprotein may be predicted on the basis of similarities shared by amino acid sequence of flaviviruses and the hepatitis C virus amino acid sequence and that a protease function resides in the amino acid sequences of flavivirus NS3 domains." *Id.* at 10-11.

a. Miyamura's Derivation of HCV-1 Subject Matter From The Inventive Entity Of The Instant Application

Applicants note that the Office Action points out that the teachings of the Miyamura patent have priority to September 15, 1989, when they appeared in the parent application 07/408,045. Office Action at 11. As noted by the Examiner, the Miyamura patent makes clear that the Figure 12 sequence, as well as the genomic organization information set out in Figure 11 and at column 7, lines 8-9, concerns HCV-1. *See* Miyamura patent, column 4, lines 33-36; column 6, line 65 to column 7, line 16. Similarly, the context surrounding the cited material at column 17, lines 16-21 of the Miyamura patent makes clear that this prediction is premised on HCV-1 sequence data. As shown below, this HCV-1 subject matter was derived by the Miyamura inventive entity from the inventive entity of the '017 patent, which is the same inventive entity as the instant application.

Applicants note that the Miyamura patent does not claim any HCV-1 sequences or methods, but rather specifically *disclaims* HCV-1. *See* Miyamura patent, column 40, line 47 to column 42, line 30, claims 1-6 (all of which contain the limitation "wherein said sequence is not homologous to the nucleotide sequence of HCV isolate HCV1"). Similarly, the Miyamura Specification makes clear that Miyamura's invention relates to J1 and J7 HCV isolates, and not to HCV-1. *See, e.g.,* Miyamura patent, column 1, lines 18-19, and column 2, line 34 to column 3, line 65.

As a further evidentiary submission, Applicants provide the Declaration of Tatsuo Miyamura Under 37 C.F.R. § 1.132 ("Miyamura Decl.", Exhibit A hereto, originally submitted

during the Reexamination of the priority patent 5,371,017), who is the first-named inventor on the Miyamura patent. Dr. Miyamura states that the Miyamura patent arose from a collaboration between himself and his colleague Dr. Izumi Saito, with Dr. Houghton and his colleagues at Chiron. Miyamura Decl. ¶5. Dr. Miyamura declares that Dr. Houghton provided him with the HCV-1 ORF sequence shown in Miyamura Figure 12 and the information regarding the HCV-1 putative genomic organization shown in Miyamura Figure 11.[6] *Id.* at ¶6. Dr. Miyamura further states that neither he nor his colleague Dr. Saito independently determined this information prior to the filing of the applications for the Miyamura patent. *Id.* Dr. Miyamura further declares that the sentence at Miyamura column 17, lines 17-21[7] "reflects work done by Dr. Houghton and his colleagues, not by Dr. Saito and myself. I believe that sentence was the contribution of Dr. Houghton." *Id.* at ¶7.

Applicants also point to the Declaration for Continuation-in-Part Application submitted to the PTO by Drs. Houghton, Choo and Kuo when the application for the '017 patent was filed, a copy of which is attached as Exhibit B hereto. In that declaration Drs. Houghton, Choo and Kuo declare that they are the "original, first and joint" inventors of the subject matter which is claimed and for which a patent is sought. Exhibit B hereto.

---

[6] Applicants note that the sequence information in Figure 12 of the Miyamura patents and the genomic organization information in Figure 11 of the Miyamura patents together provide the subject matter at Miyamura '928 patent, column 7, lines 8-9 (*i.e.*, prediction of a putative NS3 domain from about amino acid 1007 to about amino acid 1650). First, the differing numbering schemes of Figures 11 and 12 must be normalized to one another. In Figure 12, the first nucleotide of the first translated codon (part of the "putative initiator methionine") is numbered as nucleotide *320*. In contrast, the first nucleotide of the first translated codon in Figure 11 corresponds nucleotide *1*. This may be deduced, for example, because: (a) the protein encoded by the putative C domain is described as having approximately 115 amino acids (Miyamura '928 patent, column 6, line 67 to column 7, line 2); (b) the 3' boundary of the C domain in Figure 11 is designated as nucleotide 345; and (c) since a codon consists of three nucleotides, the first nucleotide of Figure 11 must represent the first nucleotide of the first translated codon (*i.e.*, 3 x 115 = 345). A nucleotide in Figure 11 can thus be correlated to a nucleotide in Figure 12 by adding 319.

The Figure 11 putative boundary numbers (which are all divisible by three) must each represent the final nucleotide of a putative domain, because nucleotide 345 corresponds to the final nucleotide of a 115-amino acid reading frame. Thus, the designation "3018 nt" indicates the final nucleotide of the final codon of NS2. Further, adding 319 to nucleotide 3019 of Figure 11 yields nucleotide 3338 of Figure 12, which corresponds to amino acid 1007. Similarly, the designation "4950 nt" indicates the last nucleotide of the last codon of the putative NS3 domain. Adding 319 to nucleotide 4950 in Figure 11 yields nucleotide 5269 in Figure 12, which corresponds to amino acid 1650. Thus, the putative amino acid range for NS3 disclosed at Miyamura '928 patent, column 7, lines 8-9, corresponds exactly to the nucleotide numbers of Miyamura Figure 11.

[7] At column 17, lines 17-21, the Miyamura '928 patent recites: "Due to the observed similarities between HCV and the Flaviviruses, deductions concerning the approximate locations of the corresponding protein domains and functions in the HCV polyprotein are possible."

b. The Legal Standard For Section 102(e) Prior Art

It has long been established that, absent a Section 102(b) statutory bar, an inventor's own work cannot be held against him as prior art under 35 U.S.C. § 102(e). Thus, for a reference patent to qualify as prior art under Section 102(e), (1) the application for the reference patent must have been by one who is legally "another" and (2) the filing date of the reference patent must be "before the invention thereof by the applicant ... ." 35 U.S.C. § 102(e). A patent cannot be relied upon as prior art under 35 U.S. C. § 102(e) when the record establishes that the relevant disclosure relied upon in the rejection is the applicant's own work, and furthermore that the relevant portions of the reference patent were obtained from the applicant. *See* MPEP § 2136.05; *In re Mathews*, 408 F.2d 1393, 161 USPQ 276 (CCPA 1969); *In re Land*, 368 F.2d 886, 151 USPQ 621 (CCPA 1966). "When the 102(e) reference patentee got knowledge of the applicant's invention from him, as by being associated with him . . . and *thereafter* describes it, he necessarily files the application *after* applicant's invention date . . . ." *Mathews*, 408 F.2d at 1396, 161 USPQ at 279 (quoting *Land*, 368 F.2d at 879, 151 USPQ at 633 (emphasis original)).

c. The Subject Matter Of Figures 11 and 12 and at Column 7, lines 8-10 and Column 17, lines 5-21 of the Miyamura Patent Is Not Citable As Prior Art

Applicants respectfully submit that the record here cannot support a conclusion that the Miyamura HCV-1 subject matter[8] relied upon in the Office Action is the invention of "another:"

- The Miyamura patents do not claim the claimed subject matter of the instant application.
- A collaboration between Dr. Houghton and Drs. Miyamura and Saito is evidenced by the recorded assignment for Serial No. 408,045 (the earliest-filed application from which Miyamura claims the benefit of filing date), which is attached as Exhibit I to this Response. In that assignment, Drs. Miyamura and Saito assign their rights in the invention to Chiron Corporation as a co-assignee with the Director General of the National Institute of Health of Japan. Further evidencing a collaboration is the fact that Michael Houghton is named as a joint inventor on U.S. patent application Serial No. 637,380, filed January 4, 1991, for the Miyamura patent. This establishes a collaboration in connection with the September 15,

[8] The subject matter at issue is the Miyamura '928 patent disclosure pertaining to the HCV-1 ORF sequence (Figure 12), the putative genomic organization of HCV-1 (Figure 11 and column 7, lines 8-9), and column 17, lines 17-21.

1989 filing date of Serial No. 408,045 (which included the subject matter relied upon in the Office Action but did not name Dr. Houghton as a joint inventor).

- Prior to the earliest filing date of Miyamura (September 15, 1989), similar or identical subject matter appeared in applications filed by the inventive entity of the instant application). *See* United States patent application Serial No. 07/355,002, filed May 18, 1989 (the '002 application")[9], Figures 62, 62.1 and 62.2 (corresponding to Miyamura Figure 12); page 45, lines 8-13 (corresponding to Miyamura patent at column 17, lines 16-21); and page 123, line 26 (corresponding approximately to Miyamura Figure 11, and Miyamura patent at column 7, lines 8-9).[10] The '002 application was incorporated by reference in the earliest-filed Miyamura application.[11] The European counterpart of the '002 application is incorporated by reference in Miyamura (*i.e.*, EP 388,232). Miyamura patent, column 5, lines 37-44.
- The declaration by the inventors of the '002 patent application, submitted when that application was filed, in which they averred that they were the "original, first and joint inventors . . . of the subject matter which is claimed and for which a patent is sought . . ." (copy attached as part of Exhibit F to this Response).
- Prior to the earliest filing date of Miyamura (September 15, 1989), HCV1 polynucleotide sequences in the HCV NS3 region appeared in applications filed by the inventive entity of the instant application. *See* Figures 32 and 47, EP 318,216 to Houghton et al.[12] The '216 application is incorporated by reference in the Miyamura patent. *See* Miyamura patent, column 5, lines 37-44.

Applicants believe that these facts prevent a conclusion that the Miyamura subject matter at issue is that of "another."

The Miyamura Declaration, coupled with the declaration by the inventors of the '017 patent (and the instant application) that was filed with the application for the '017 patent, further supports that the Miyamura HCV-1 disclosure relied upon by the Examiner was obtained by the

---

[9] This application was incorporated by reference into Serial No. 07/456,637 (Exhibit D hereto, page 1, lines 10-12; page 2, lines 2-3), which itself was incorporated by reference into the '017 patent (*see, e.g.*, '017 patent, column 2, lines 43-49).

[10] For the convenience of the Examiner a copy of the following from the '002 application is attached as Exhibit F to this Response: Figure 62, 62-1 and 62-2, and pages 17 (describing Figure 62), 45 and 123. Also attached as Exhibit F to this Response is a copy of the Filing Receipt for the '002 application, showing the inventorship.

[11] *See* page 8, lines 3-4 of the 'Miyamura '045 application, attached as Exhibit G to this Response.

Miyamura inventors from the inventors of the instant application (through Dr. Houghton), and that this disclosure was the own work of the inventors of the instant application. *See Mathews*; Exhibits A & B hereto. Accordingly, the portions of the Miyamura patent pertaining to the HCV-1 ORF sequence (Figure 12), putative genomic organization of HCV-1 (Figure 11 and column 7, lines 8-9 of the Miyamura patent), and column 17, lines 17-21 of the Miyamura patent, have effectively been removed as prior art under 35 U.S.C. § 102(e).

3. The Disclosure in Miyamura of Methods of Production of Polypeptides Encoded by the HCV Genome, Taken Alone, Does Not Render Claims 27-35 Obvious

The Office Action also relies upon "Examples I-IV at columns 28-39" to contend that Miyamura et al. "teach preparation of cloning vectors, and transformed host cells comprising the vectors, comprising inserts of specific, defined, regions found anywhere in a nucleic acid sequence encoding all or part of an hepatitis C virus polyprotein." Office Action at 11. The patent "explicitly teach, at cols. 8-10, that expression vectors comprising transcriptional and translational regulatory elements operably linked to a polynucleotide encoding a desired regions [sic] of the HCV polyprotein should be used to produce desired portions of the hepatitis C virus polyprotein in host cells." *Id.* Miyamura et al. "further suggest preparation of expression constructs providing fusions of hepatitis C virus amino acid sequence regions with proteins commonly used in the art as fusion partners such as β-galactosidase and superoxide dismutase [SOD]." Office Action at 11 (bracketed text in original).

As shown above, there is no HCV protease sequence in Miyamura that is citable as prior art against the instant application. Absent a prior art HCV protease sequence to express, disclosure concerning standard methodology for expressing polypeptides cannot by itself support a *prima facie* obviousness rejection.

4. The Rejection Must Be Withdrawn Based on the Removal of the Subject Matter of Figures 11 and 12, Column 17, lines 17-21, and Column 7, lines 8-9, of the Miyamura Patent.

The Miyamura patent cannot support a *prima facie* obviousness rejection because, as shown above, the key disclosures thereof are not available as prior art. Applicants submit that

---

[12] For the convenience of the Examiner a copy of the following from the '216 application is attached as Exhibit J to this Response: Page 1, page 9 (describing Figures 32 and 47), and Figures 32-1 to 32-7 and 47-1 to 47-8.

the outstanding rejection must be withdrawn in the absence of the subject matter derived from the inventors of the instant application. The secondary references on which the Examiner relies -- Gorbalenya, Bazan and Miller -- do not remedy Miyamura's deficiencies.

Gorbalenya and Bazan concern flaviviruses. Neither reference mentions HCV, and HCV is not a flavivirus.[13] Miller discloses only nucleic acid sequence of a helicase region within the NS3 domain of HCV. Miller does not disclose any sequence encoding a protease. Thus, there is no link between Gorbalenya, Bazan or Miller to the pending claims of the instant application. In view of the foregoing, none of the secondary references cited by the Examiner can support a *prima facie* obviousness rejection.[14]

[13] Miyamura itself states that HCV is "a new viral class" distinct from flaviviruses. *See, e.g.*, Miyamura '928 patent, column 2, lines 1-8; *see also* C. Rice, "Flaviviridae -- The Viruses And Their Replication", in Fields Virology, Vol. 1 (B. Fields) (3d ed. 1995), pages 932-33, which is attached as Exhibit C to this Response.

[14] For the record, Applicants do not concede that the claims are *prima facie* obvious over Miyamura in view of Gorbalenya, Bazan and Miller even if the HCV-1 subject matter of Miyamura is legally available prior art (which it is not).

CONCLUSION

In light of the arguments set forth above, Applicants earnestly believe that they are entitled to a letters patent, and respectfully solicit the Examiner to expedite prosecution of this patent application to issuance. If it is determined that a telephone conference would expedite the prosecution of this application, the Examiner is invited to telephone the undersigned at the number given below.

In the event the U.S. Patent and Trademark office determines that an extension and/or other relief is required, applicant petitions for any required relief including extensions of time and authorizes the Commissioner to charge the cost of such petitions and/or other fees due in connection with the filing of this document to Deposit Account No. 03-1952 referencing docket no. 223002010004. However, the Commissioner is not authorized to charge the cost of the issue fee to the Deposit Account.

Dated: December 30, 2004

Respectfully submitted,

By _____

Shantanu Basu
Registration No.: 43,318
MORRISON & FOERSTER LLP
755 Page Mill Road
Palo Alto, California 94304
(650) 813-5995

PATENT
Docket No. 223002010030

CERTIFICATE OF MAILING BY "EXPRESS MAIL"

Express Mail Label No.: EL_____ Date of Deposit: February __, 2000

I hereby certify that this paper or fee is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 C.F.R. § 1.10 on the date indicated above and is addressed to: Assistant Commissioner for Patents, Washington, D.C. 20231

_____

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Reexamination Control No. 90/005,512

In the Reexamination of:

U.S. Patent No.: 5,371,017

Inventors: Michael HOUGHTON,
Qui-Lim CHOO, and
George KUO

Issue Date: December 6, 1994

Group Art Unit: 1652

For: HEPATITIS C VIRUS PROTEASE

Examiner: Rebecca Prouty

pa-561310

PATENT
Docket No. 223002010030
Reexam. Control No. 90/005,512

DECLARATION OF TATSUO MIYAMURA UNDER 37 C.F.R. § 1.132

Assistant Commissioner for Patents
Washington, D.C. 20231

I, Tatsuo Miyamura, declare as follows:

1. I am a Japanese citizen currently residing at 4-21-22-113, Hamadayama Suginami-ku, Tokyo 168-0065, Japan.

2. I am a scientist at the National Institute of Health of Japan ("the present name National Institute of Infectious Diseases") and have been employed by there for 30 years. I was awarded a medicine from Niigata University in 1970. I have worked from Niigata University in 1976. in the field of virology for over 30 years, and have many publications in that field.

3. I am the lead inventor of Miyamura et al. United States patent No. 5,372,928 ("the Miyamura patent"). Dr. Michael Houghton of Chiron Corporation ("Chiron") and six others are co-inventors with me on that patent.

4. The Miyamura patent issued from a series of applications, the earliest of which was filed September 15, 1989. Chiron patent attorneys were involved in writing the applications for the Miyamura patent.

5. The Miyamura patent arose from a collaboration between myself and my colleague Keizo Sato at the Institute with Dr. Houghton and his colleagues at Chiron.

6. As part of the collaboration, Dr. Houghton gave me at least the following:

- the nucleotide sequence for the open reading frame of Hepatitis C Virus 1 ("HCV-1"), which is shown in Figure 12 of the Miyamura patent; and
- information regarding the putative genomic organization of HCV-1, which is shown in Figure 11 of the Miyamura patent.

Prior to the filing of the applications for the Miyamura patent, neither Dr. Saito nor I had independently determined this information.

At column 17, lines 17-21 of the Miyamura patent, it states that "[d]ue to the observed similarities between HCV and the Flaviviruses, deductions concerning the approximate locations of the corresponding protein domains and functions in the HCV polyprotein are possible." That sentence reflects work done by Dr. Houghton and his colleagues and by Dr. Saito and myself. I believe that sentence was the contribution of Dr. Houghton.

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application, or any patent issued thereon.

Signed this 9 th day of February, 2001, at Tokyo, Japan.

Tatsuo Miyamura

Dr. Tatsuo Miyamura

Houghton. I will send C.V.

07/680296

CHIRON CORPORATION
4560 Horton Street, Emeryville, CA 94608
(415) 655-8730

Honorable Commissioner of Patents
and Trademarks
Washington, D.C. 20231

Attorney's Docket No. 0100.002

SIR:

NEW APPLICATION TRANSMITTAL

Transmitted herewith for filing is the patent application of Inventor(s): Michael Houghton, Qui-Lim Choo and George Kuo

Title: HEPATITIS C VIRUS PROTEASE

CERTIFICATION UNDER 37 CFR §1.10

I hereby certify that this New Application and the documents referred to as enclosed herein are being deposited with the United States Postal Service on this date 4 April 1991 in an envelope bearing "Express Mail Post Office To Addressee" Mailing Label Number RB561501127 addressed to: Patent Application, Honorable Commissioner of Patents and Trademarks, Washington, D.C. 20231.

April 4, 1991
Gyne Riser

Gyne M. Riser
Signature

Enclosed are:

1. The papers required for filing date under CFR §1.53(b):
   47 Pages of specification (including claims);
   23 Sheets of drawings.
   ___ formal
   X informal
2. X Combined Declaration/Power of Attorney (unsigned)
3. ___ Assignment of the invention to Chiron Corporation
5. Fee Calculation
   ___ Amendment changing number of claims or deleting multiple dependencies is enclosed.

CLAIMS AS FILED

	Number filed	Number extra	Rate	Basic Fee $630.00
Total Claims	26 - 20 =	6	X $20.00	$120.00
Independent Claims	6 - 3 =	3	X $60.00	$180.00
0 Multiple dependent claim(s), if any		0	$120.00	0

*If less than zero, enter "0".

Filing Fee Calculation. . . . . $ 930.00

6. ___ Small Entity Statement - verified statement enclosed
   50% Filing Fee Reduction (if applicable). . $ -----
7. Other Fees
   ___ Recording Assignment ($8.00)
   0 Other fees . . . . . $ -----
   Specify _____ . . . . . $ -----

Total Fees Enclosed . . . . . $ 930.00

8. Payment of Fees
   X Check(s) in the amount of $ 930.00 enclosed.
   ___ Charge Account No. 03-1664 in the amount of $ 0 . A duplicate of this transmittal is attached.
9. X Authorization to Charge Additional Fees
   The Commissioner is hereby authorized to charge any additional fees (or credit any overpayment) associated with this communication and which may be required under 37 CFR §1.16 or §1.17 to Account No. 03-1664. A duplicate sheet is attached.
10. ___ Information Disclosure Statement
11. X Return Receipt Postcard
12. ___ Other: Specify _____

Dated 4/4/91

By: Grant D. Green

Name: Grant D. Green
Registration No. 31,259

Paper # 3

Patent
Atty. Docket No. 0100.002

DECLARATION FOR CONTINUATION-IN-PART-APPLICATION

As a below named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below next to my name.

I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled

HEPATITIS C VIRUS PROTEASE

the specification of which (check one) _____ is attached hereto __X__ was filed on April 4, 1991 as Application Serial No. 07/680,296 and was amended on ____ (if applicable).

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims, as amended referred to above.

I acknowledge the duty to disclose information which is material to the examination of this application in accordance with Title 37, Code of Federal Regulations, Section 1.56(a).

==========

I hereby claim foreign priority benefits under Title 35, United States Code, §119 of any foreign application(s) for patent or inventor's certificate listed below and have also identified below any foreign application for patent or inventor's certificate having a filing date before that of the application on which priority is claimed:

Prior Foreign Application(s) Number	Country	Day/Month/Year Filed	Priority Claimed Yes	No

==========

I hereby claim the benefit under Title 35, United States Code, §120 of any United States application(s) listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States application in the manner provided by the first paragraph of Title 35, United States Code, §112, I acknowledge the duty to disclose material information as defined in Title 37, Code of Federal Regulations, §1.56(a) which occurred between the filing date of the prior application and the national or PCT international filing date of this application:

Application Serial No.	Filing Date	Status Patented, Pending, Abandoned
07/505,433	04/04/90	Pending

==========

I hereby declare that all statements herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements and the like so made are punishable by fine or imprisonment or both, under Section 1001 of Title 18 of the United States code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

=======

Full name of sole or first inventor Michael Houghton 4010

Inventor's signature MHoughton Date 6/17/91

Residence Danville, California CA

Citizenship United Kingdom

Post Office Address 53 Rosemead Court, Danville, CA ~~94526~~ 94526 MH

FIELDS

VIROLOGY

Third Edition

Volume 1

Editors-in-Chief

Bernard N. Fields, M.D.
Chairman, Department of Microbiology and Molecular Genetics
Harvard Medical School
Boston, Massachusetts

David M. Knipe, Ph.D.
Department of Microbiology and Molecular Genetics
Harvard Medical School
Boston, Massachusetts

Peter M. Howley, M.D.
Chairman, Department of Pathology
Harvard Medical School
Boston, Massachusetts

Associate Editors

Robert M. Chanock, M.D.
Laboratory of Infectious Diseases
National Institute of Allergy and Infectious Disease
National Institutes of Health
Bethesda, Maryland

Joseph L. Melnick, Ph.D., Sc.D.
Department of Virology and Epidemiology
Baylor College of Medicine
Texas Medical Center
Houston, Texas

Thomas P. Monath, M.D.
Ora Vax, Inc.
Cambridge, Massachusetts

Bernard Roizman, Sc.D.
Department of Molecular Genetics and Cell Biology
University of Chicago
Chicago, Illinois

Stephen E. Straus, M.D.
National Institute of Allergy and Infectious Diseases
National Institutes of Health
Bethesda, Maryland

Lippincott - Raven
PUBLISHERS
Philadelphia • New York

Lippincott-Raven Publishers, Philadelphia, 227 East Washington Square, Philadelphia, PA 19106

Made in the United States of America

Library of Congress Cataloging-in-Publication Data

Fields Virology / editors-in-chief, Bernard N. Fields, David M. Knipe, Peter M. Howley ; associate editors, Robert M. Chanock . . . [et al.].—3rd ed.
p. cm.
Includes bibliographical references and index.
ISBN 0-7817-0253-4 (alk. paper)
1. Virology. I. Fields, Bernard N. II. Knipe, David M. (David Mahan), III. Howley, Peter M.
[DNLM: 1. Viruses. 2. Virus Diseases. QW 160 F463 1995]
QR360.V5125 1995
616'.0194—dc20
DNLM/DLC
For Library of Congress 95-12782

The material contained in this volume was submitted as previously unpublished material, except in the instances in which credit has been given to the source from which some of the illustrative material was derived.

Great care has been taken to maintain the accuracy of the information contained in the volume. However, neither Lippincott-Raven Publishers nor the editors can be held responsible for errors or for any consequences arising from the use of the information contained herein.

Materials appearing in this book prepared by individuals as part of their official duties as U.S. Government employees are not covered by the above-mentioned copyright.

9 8 7 6 5 4 3 2 1

CLASSIFICATION

Flaviviruses

The *Flavivirus* genus includes more than 68 members separated into groups on the basis of serological relatedness (37) (Table 1; see Chapter 31). More recently, similar relationships have been found by comparison of flavivirus genome sequences (21,171). Most flaviviruses are arthropod-borne, being transmitted to vertebrates by chronically infected mosquito or tick vectors. However, isolates from bats and rodents, without known insect vectors, also have been identified. Arthropod-borne flaviviruses cause significant human and animal disease and are distributed worldwide (206,289) (see also Chapter 31). Clinical symptoms vary and include fever, encephalitis and hemorrhagic fever (see Chapter 31). Entities of major global concern include dengue fever with its associated dengue hemorrhagic fever (DHF) and shock syndrome (DSS) (113,114), Japanese encephalitis (JE) (207), and YF. Tick-borne encephalitis (TBE), Kyasanur Forest disease, West Nile encephalitis (WN), St. Louis encephalitis (SLE), and Murray Valley encephalitis (MVE) are other important agents of regional endemic or epidemic disease (206) (see Chapter 31). Thus far, vaccination is available for YF, using the live-attenuated 17D strain (318), and for TBE and JE using inactivated virus (124).

Pestiviruses

Currently recognized pestiviruses include three serologically related animal pathogens (203) (see Chapter 33). These include the type virus, bovine viral diarrhea virus (BVDV), classical swine fever virus (CSFV; also called hog cholera virus), and border disease virus (BDV) of sheep. The border disease group has recently been shown to comprise BVDV-like isolates as well as true BDV strains (16). Pestivirus diseases are widespread and of major economic importance to the livestock industry (203). Transmission occurs by direct or indirect contact as well as by congenital routes. Clinical manifestations vary and include inapparent infections, acute or persistent subclinical infections, fetal death and congenital abnormalities, wasting disease, and an acute fatal illness called mucosal disease (MD) (203). Recently, a new variant of BVDV has been identified that causes severe thrombocytopenia and hemorrhagic syndrome in adult animals (22,63,64,255). Live-

TABLE 1. *Members of the Flaviviridae*

	Group	Type member
Flaviviruses	Tick-borne encephalitis (12[a], T[b])	Central European encephalitis (TBE-W)
		Far Eastern encephalitis (TBE-FE)
	Rio Bravo[c] (6, T)	Rio Bravo
	Japanese encephalitis (10, M)	Japanese encephalitis (JE)
		Kunjin (KUN)
		Murray Valley encephalitis (MVE)
		St. Louis encephalitis (SLE)
		West Nile (WN)
	Tyuleniy (3, T)	Tyuleniy
	Ntaya[c] (5, M)	Ntaya
	Uganda S (4, M)	Uganda S
	Dengue (4, M)	Dengue type 1 (DEN1)
		Dengue type 2 (DEN2)
		Dengue type 3 (DEN3)
		Dengue type 4 (DEN4)
	Modoc (5, U)	Modoc
	Ungrouped[c] (17, M)	Yellow fever (YF)
Pestiviruses	Bovine viral diarrhea	Bovine viral diarrhea (BVDV)
	Classical swine fever	Hog cholera or classical swine fever (CSFV[e])
	Border disease	Border disease (BDV)
Hepatitis C viruses[d]	Hepatitis C	Hepatitis C (HCV)

[a]Number of recognized members in each antigenic group [from Calisher et al. (37)].

[b]Arthropod vectors: T, tick; M, mosquito; U, unidentified or no vector.

[c]Arthropod vectors for some members of these groups have not been identified. The ungrouped flaviviruses include mosquito- and tick-transmitted viruses as well as some with no known vector.

[d]The hepatitis C viruses include a large number of isolates that can be divided into several groups or genotypes on the basis of genetic divergence (36,269,291). An official name for this genus and a standardized nomenclature for different genotypes have not yet been agreed upon.

[e]In the pestivirus literature, HCV has been a common abbreviation for hog cholera virus. More recent publications and this chapter use CSFV to avoid confusion with the human hepatitis C viruses.

attenuated strains and inactivated virus preparations are available for vaccination against CSFV and BVDV (203), but there is need for improved pestivirus vaccines (see Chapter 33).

Hepatitis C Viruses

The hepatitis C viruses (HCV) compose the remaining genus of the Flaviviridae. After the development of diagnostic tests for hepatitis A virus (Chapter 24) and hepatitis B virus (Chapter 86), an additional agent, which could be experimentally transmitted to chimpanzees (4,139,309), became recognized as the major cause of transfusion-acquired hepatitis. The causative agent, previously designated non-A, non-B hepatitis virus and now referred to as HCV, was identified in 1989 (54). Development of diagnostic tests to identify HCV carriers among blood donors (52,162) has already markedly reduced the frequency of posttransfusion hepatitis (3). Humans are the only known natural host for HCV; there is no evidence for vector-mediated transmission. HCV infection is found throughout the world, and the prevalence of anti-HCV antibodies ranges from 0.4% to 2% in most developed countries to more than 14% in Egypt (129) (see Chapter 32). Besides transmission via blood or blood products, or less frequently by sexual and congenital routes, sporadic cases occur that are not associated with known risk factors and account for more than 40% of HCV cases (5,194). Infections are usually chronic (6), and clinical outcomes (138) (see Chapter 32) range from an inapparent carrier state to acute hepatitis, chronic active hepatitis, and cirrhosis, which is strongly associated with the development of hepatocellular carcinoma (HCC) (288). Although alpha interferon has been shown to be useful for the treatment of some patients with chronic HCV infections (65,71) and subunit vaccines show some promise in the chimpanzee model (53), future efforts are needed to develop more effective therapies and vaccines. The considerable diversity observed among different HCV isolates (36,290), the emergence of genetic variants in chronically infected individuals (76,131,151,152, 163,170,227,336,337), and the lack of protective immunity elicited after HCV infection (81,245) present major challenges toward these goals. –

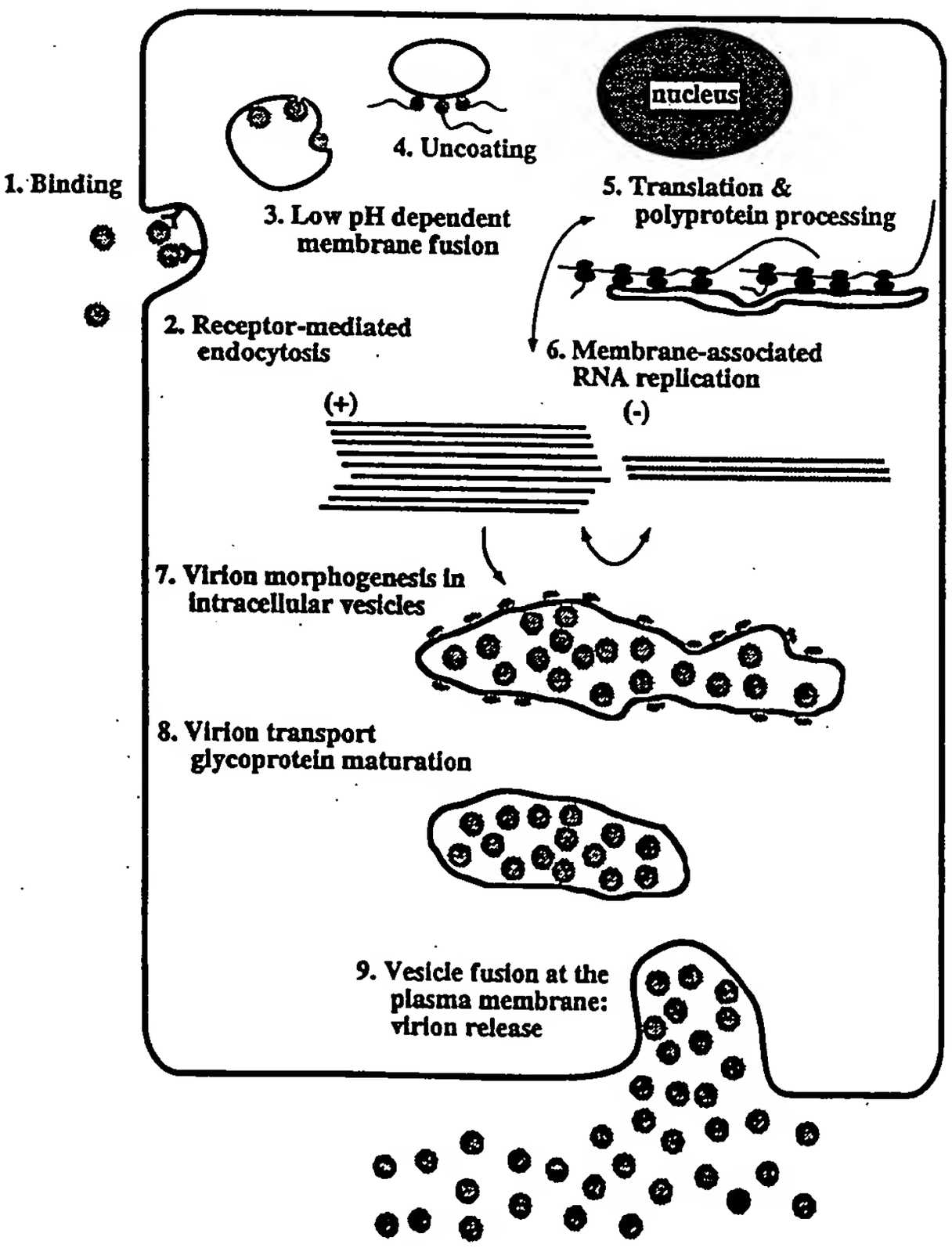


FIG. 1. The flavivirus lifecycle.

CERTIFICATE OF MAILING BY "EXPRESS MAIL"
"Express Mail" Mailing Label No. 329063835
Date of Deposit 21 DEC 1989
I hereby certify that this paper or fee is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 CFR 1.10 on the date indicated above and is addressed to the Commissioner of Patents and Trademarks, Washington, D.C. 20231.

Gladys H. Monroy
(Typed or Printed Name of Person Mailing Paper or Fee)

Gladys H. Monroy
(Signature of Person Mailing Paper or Fee)

NANBV DIAGNOSTICS AND VACCINES

Cross-Reference to Related Applications

This application is a continuation-in-part of attorney docket 2300-0063.28 (U.S.S.N. 07/355,002) filed 18 May 1989; which is a continuation-in-part of attorney docket number 2300-0063.29 (U.S.S.N. 07/341,334) filed 20 April 1989; which is a continuation-in-part of attorney docket number 2300-0063.59 (PCT/US88/04125) filed 18 November 1988, converted to U.S. National phase on 21 April 1989 and assigned attorney docket number 2300-0063.26 (U.S.S.N. 353,896); and a continuation-in-part of attorney docket number 2300-0063.25 (U.S.S.N. 07/325,338) filed 17 March 1989 (now abandoned); which are continuations-in-part of attorney docket number 2300-0063.24 (U.S.S.N. 271,450) filed 14 November 1988, now abandoned; which is a continuation-in-part of attorney docket number 2300-0063.23 (U.S.S.N. 263,584) filed 26 October 1988, now abandoned; which is a continuation-in-part of attorney docket number 2300-0063.22 (formerly 2300-0237, U.S.S.N. 191,263) filed 6 May 1988, now abandoned; which is a continuation-in-part of attorney docket number 2300-0063.21 (formerly 2300-0228, U.S.S.N. 161,072) filed 26 February 1988, now abandoned; which is a continuation-in-part of attorney docket number 2300-0063.20 (formerly 2300-0219, U.S.S.N. 139,886) filed 30 December 1987, now abandoned; which is a continuation-in-part of attorney docket number 2300-0063 (formerly 2300-

0203, U.S.S.N. 122,714) filed 18 November 1987, now abandoned; the aforementioned applications are, in their entirety, incorporated herein by reference.

Technical Field

The invention relates to materials and methodologies for managing the spread of non-A, non-B hepatitis virus (NANBV) infection. More specifically, it relates to diagnostic DNA fragments, diagnostic proteins, diagnostic antibodies and protective antigens and antibodies for an etiologic agent of NANB hepatitis, i.e., hepatitis C virus.

References Cited in the Application

Barr et al. (1986), Biotechniques 4:428.
Bradley et al. (1985), Gastroenterology 88:773.
Botstein (1979), Gene 8:17.
Brinton, M.A. (1986) in THE VIRUSES: THE TOGAVIRIDAE AND FLAVIVIRIDAE (Series eds. Fraenkel-Conrat and Wagner, vol. eds. Schlesinger and Schlesinger, Plenum Press), p.327-374.
Broach (1981) in: Molecular Biology of the Yeast Saccharomyces, Vol. 1, p.445, Cold Spring Harbor Press.
Broach et al. (1983), Meth. Enz. 101:307.
Catty (1988), ANTIBODIES, Volume 1: A Practical Approach (IRL Press).
Chaney et al. (1986), Cell and Molecular Genetics 12:237.
Chakrabarti et al. (1985), Mol. Cell Biol. 5:3403.
Chang et al. (1977), Nature 198:1056.
Chen and Seeburg (1985), DNA 4:165.
Chirgwin et al. (1979), Biochemistry 18:5294.
Chomczynski and Sacchi (1987), Analytical Biochemistry 162:156.
Choo et al. (1989), Science 244359.
Clewell et al. (1969), Proc. Natl. Acad. Sci. USA 62:1159.
Clewell (1972), J. Bacteriol. 110:667.

encoded in the major ORF of the HCV genome. Also indicated in the figure are the possible functions of the flaviviral polypeptides cleaved from the flaviral polyprotein. In addition, the relative placements of the HCV polypeptides, $NANB_{5-1-1}$ and C100, with respect to the putative HCV polyprotein are indicated.

Fig. 70 shows relevant characteristics of AcNPV transfer vectors used for high level expression of nonfused foreign proteins. It also shows a restriction endonuclease map of the transfer vector pAc373.

Fig. 71 shows the nucleotide sequence of clone 6k, the part of the sequence which overlaps clone 16jh, and the amino acids encoded therein.

Fig. 72 shows a composite cDNA sequence derived from overlapping clones clones b114a, 18g, ag30a, CA205a, CA290a, CA216a, pi14a, CA167b, CA156e, CA84a, CA59a, K9-1 (also called k9-1), 26j, 13i, 12f, 14i, 11b, 7f, 7e, 8h, 33c, 40b, 37b, 35, 36, 81, 32, 33b, 25c, 14c, 8f, 33f, 33g, 39c, 35f, 19g, 26g, 15e, b5a, 16jh and 6k; also shown are the amino acids encoded in the positive strand of the cDNA (which is the equivalent of the HCV RNA).

Fig. 73 shows the linkers used in the construction of $pS3\text{-}56_{C100m}$.

Fig. 74 shows the nucleotide sequence of the HCV cDNA in clone 31, the amino acids encoded therein, and putative restriction enzyme sites encoded therein.

Fig. 75 shows the nucleotide sequence of the HCV cDNA in clone p131jh, and its overlap with the nucleotide sequence in clone 6k.

Fig. 76 shows a flow chart for construction of the expression vector $pC100^{-}d\#3$.

Fig. 77 shows a flow chart for construction of the expression vector pS2d#9.

Fig. 78 shows a flow chart for construction of the expression vector pNS11d/13.

FIGURE 72-3

```
                CTTCGACGGACGTTGACCTGCGCCCCGCTTGCAACGCTAGACCTTCTGTCCCTGTCCAGG

                GluLeuSerProLeuLeuLeuThrThrThrGlnTrpGlnValLeuProCysSerPheThr
1981            GAGCTCAGCCCGTTACTGCTGACCACTACACAGTGGCAGGTCCTCCCGTGTTCCTTCACA
                CTCGAGTCGGGCAATGACGACTGGTGATGTGTCACCGTCCAGGAGGGCACAAGGAAGTGT

                ThrLeuProAlaLeuSerThrGlyLeuIleHisLeuHisGlnAsnIleValAspValGln
2041            ACCCTACCAGCCTTGTCCACCGGCCTCATCCACCTCCACCAGAACATTGTGGACGTGCAG
                TGGGATGGTCGGAACAGGTGGCCGGAGTAGGTGGAGGTGGTCTTGTAACACCTGCACGTC

                TyrLeuTyrGlyValGlySerSerIleAlaSerTrpAlaIleLysTrpGluTyrValVal
2101            TACTTGTACGGGGTGGGGTCAAGCATCGCGTCCTGGGCCATTAAGTGGGAGTACGTCGTT
                ATGAACATGCCCCACCCCAGTTCGTAGCGCAGGACCCGGTAATTCACCCTCATGCAGCAA

                LeuLeuPheLeuLeuLeuAlaAspAlaArgValCysSerCysLeuTrpMetMetLeuLeu
2161            CTCCTGTTCCTTCTGCTTGCAGACGCGCGCGTCTGCTCCTGCTTGTGGATGATGCTACTC
                GAGGACAAGGAAGACGAACGTCTGCGCGCGCAGACGAGGACGAACACCTACTACGATGAG

                IleSerGlnAlaGluAlaAlaLeuGluAsnLeuValIleLeuAsnAlaAlaSerLeuAla
2221            ATATCCCAAGCGGAGGCGGCTTTGGAGAACCTCGTAATACTTAATGCAGCATCCCTGGCC
                TATAGGGTTCGCCTCCGCCGAAACCTCTTGGAGCATTATGAATTACGTCGTAGGGACCGG

                GlyThrHisGlyLeuValSerPheLeuValPhePheCysPheAlaTrpTyrLeuLysGly
2281            GGGACGCACGGTCTTGTATCCTTCCTCGTGTTCTTCTGCTTTGCATGGTATTTGAAGGGT
                CCCTGCGTGCCAGAACATAGGAAGGAGCACAAGAAGACGAAACGTACCATAAACTTCCCA

                LysTrpValProGlyAlaValTyrThrPheTyrGlyMetTrpProLeuLeuLeuLeuLeu
2341            AAGTGGGTGCCCGGAGCGGTCTACACCTTCTACGGGATGTGGCCTCTCCTCCTGCTCCTG
                TTCACCCACGGGCCTCGCCAGATGTGGAAGATGCCCTACACCGGAGAGGAGGACGAGGAC

                LeuAlaLeuProGlnArgAlaTyrAlaLeuAspThrGluValAlaAlaSerCysGlyGly
2401            TTGGCGTTGCCCCAGCGGGCGTACGCGCTGGACACGGAGGTGGCCGCGTCGTGTGGCGGT
                AACCGCAACGGGGTCGCCCGCATGCGCGACCTGTGCCTCCACCGGCGCAGCACACCGCCA

                ValValLeuValGlyLeuMetAlaLeuThrLeuSerProTyrTyrLysArgTyrIleSer
2461            GTTGTTCTCGTCGGGTTGATGGCGCTGACTCTGTCACCATATTACAAGCGCTATATCAGC
                CAACAAGAGCAGCCCAACTACCGCGACTGAGACAGTGGTATAATGTTCGCGATATAGTCG

                TrpCysLeuTrpTrpLeuGlnTyrPheLeuThrArgValGluAlaGlnLeuHisValTrp
2521            TGGTGCTTGTGGTGGCTTCAGTATTTTCTGACCAGAGTGGAAGCGCAACTGCACGTGTGG
                ACCACGAACACCACCGAAGTCATAAAAGACTGGTCTCACCTTCGCGTTGACGTGCACACC

                IleProProLeuAsnValArgGlyGlyArgAspAlaValIleLeuLeuMetCysAlaVal
2581            ATTCCCCCCCCTCAACGTCCGAGGGGGGCGCGACGCCGTCATCTTACTCATGTGTGCTGTA
                TAAGGGGGGGAGTTGCAGGCTCCCCCCGCGCTGCGGCAGTAGAATGAGTACACACGACAT

                HisProThrLeuValPheAspIleThrLysLeuLeuLeuAlaValPheGlyProLeuTrp
2641            CACCCGACTCTGGTATTTGACATCACCAAATTGCTGCTGGCCGTCTTCGGACCCCTTTGG
                GTGGGCTGAGACCATAAACTGTAGTGGTTTAACGACGACCGGCAGAAGCCTGGGGAAACC

                IleLeuGlnAlaSerLeuLeuLysValProTyrPheValArgValGlnGlyLeuLeuArg
2701            ATTCTTCAAGCCAGTTTGCTTAAAGTACCCTACTTTGTGCGCGTCCAAGGCCTTCTCCGG
                TAAGAAGTTCGGTCAAACGAATTTCATGGGATGAAACACGCGCAGGTTCCGGAAGAGGCC

                PheCysAlaLeuAlaArgLysMetIleGlyGlyHisTyrValGlnMetValIleIleLys
2761            TTCTGCGCGTTAGCGCGGAAGATGATCGGAGGCCATTACGTGCAAATGGTCATCATTAAG
                AAGACGCGCAATCGCGCCTTCTACTAGCCTCCGGTAATGCACGTTTACCAGTAGTAATTC

                LeuGlyAlaLeuThrGlyThrTyrValTyrAsnHisLeuThrProLeuArgAspTrpAla
2821            TTAGGGGCGCTTACTGGCACCTATGTTTATAACCATCTCACTCCTCTTCGGGACTGGGCG
                AATCCCCGCGAATGACCGTGGATACAAATATTGGTAGAGTGAGGAGAAGCCCTGACCCGC

                HisAsnGlyLeuArgAspLeuAlaValAlaValGluProValValPheSerGlnMetGlu
2881            CACAACGGCTTGCGAGATCTGGCCGTGGCTGTAGAGCCAGTCGTCTTCTCCCAAATGGAG
                GTGTTGCCGAACGCTCTAGACCGGCACCGACATCTCGGTCAGCAGAAGAGGGTTTACCTC
```

FIGURE 72-4

```
      ThrLysLeuIleThrTrpGlyAlaAspThrAlaAlaCysGlyAspIleIleAsnGlyLeu
2941  ACCAAGCTCATCACGTGGGGGGCAGATACCGCCGCGTGCGGTGACATCATCAACGGCTTG
      TGGTTCGAGTAGTGCACCCCCCGTCTATGGCGGCGCACGCCACTGTAGTAGTTGCCGAAC

      ProValSerAlaArgArgGlyArgGluIleLeuLeuGlyProAlaAspGlyMetValSer
3001  CCTGTTTCCGCCCGCAGGGGCCGGGAGATACTGCTCGGGCCAGCCGATGGAATGGTCTCC
      GGACAAAGGCGGGCGTCCCCGGCCCTCTATGACGAGCCCGGTCGGCTACCTTACCAGAGG

      LysGlyTrpArgLeuLeuAlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeu
3061  AAGGGGTGGAGGTTGCTGGCGCCCATCACGGCGTACGCCCAGCAGACAAGGGGCCTCCTA
      TTCCCCACCTCCAACGACCGCGGGTAGTGCCGCATGCGGGTCGTCTGTTCCCCGGAGGAT

      GlyCysIleIleThrSerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGln
3121  GGGTGCATAATCACCAGCCTAACTGGCCGGGACAAAAACCAAGTGGAGGGTGAGGTCCAG
      CCCACGTATTAGTGGTCGGATTGACCGGCCCTGTTTTTGGTTCACCTCCCACTCCAGGTC

      IleValSerThrAlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThr
3181  ATTGTGTCAACTGCTGCCCAAACCTTCCTGGCAACGTGCATCAATGGGGTGTGCTGGACT
      TAACACAGTTGACGACGGGTTTGGAAGGACCGTTGCACGTAGTTACCCCACACGACCTGA

      ValTyrHisGlyAlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMet
3241  GTCTACCACGGGGCCGGAACGAGGACCATCGCGTCACCCAAGGGTCCTGTCATCCAGATG
      CAGATGGTGCCCCGGCCTTGCTCCTGGTAGCGCAGTGGGTTCCCAGGACAGTAGGTCTAC

      TyrThrAsnValAspGlnAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeu
3301  TATACCAATGTAGACCAAGACCTTGTGGGCTGGCCCGCTCCGCAAGGTAGCCGCTCATTG
      ATATGGTTACATCTGGTTCTGGAACACCCGACCGGGCGAGGCGTTCCATCGGCGAGTAAC

      ThrProCysThrCysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIle
3361  ACACCCTGCACTTGCGGCTCCTCGGACCTTTACCTGGTCACGAGGCACGCCGATGTCATT
      TGTGGGACGTGAACGCCGAGGAGCCTGGAAATGGACCAGTGCTCCGTGCGGCTACAGTAA

      ProValArgArgArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyr
3421  CCCGTGCGCCGGCGGGGTGATAGCAGGGGCAGCCTGCTGTCGCCCCGGCCCATTTCCTAC
      GGGCACGCGGCCGCCCCACTATCGTCCCCGTCGGACGACAGCGGGGCCGGGTAAAGGATG

      LeuLysGlySerSerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePhe
3481  TTGAAAGGCTCCTCGGGGGGTCCGCTGTTGTGCCCCGCGGGGCACGCCGTGGGCATATTT
      AACTTTCCGAGGAGCCCCCCAGGCGACAACACGGGGCGCCCCGTGCGGCACCCGTATAAA

      ArgAlaAlaValCysThrArgGlyValAlaLysAlaValAspPheIleProValGluAsn
3541  AGGGCCGCGGTGTGCACCCGTGGAGTGGCTAAGGCGGTGGACTTTATCCCTGTGGAGAAC
      TCCCGGCGCCACACGTGGGCACCTCACCGATTCCGCCACCTGAAATAGGGACACCTCTTG

      LeuGluThrThrMetArgSerProValPheThrAspAsnSerSerProProValValPro
3601  CTAGAGACAACCATGAGGTCCCCGGTGTTCACGGATAACTCCTCTCCACCAGTAGTGCCC
      GATCTCTGTTGGTACTCCAGGGGCCACAAGTGCCTATTGAGGAGAGGTGGTCATCACGGG

      GlnSerPheGlnValAlaHisLeuHisAlaProThrGlySerGlyLysSerThrLysVal
3661  CAGAGCTTCCAGGTGGCTCACCTCCATGCTCCCACAGGCAGCGGCAAAAGCACCAAGGTC
      GTCTCGAAGGTCCACCGAGTGGAGGTACGAGGGTGTCCGTCGCCGTTTTCGTGGTTCCAG

      ProAlaAlaTyrAlaAlaGlnGlyTyrLysValLeuValLeuAsnProSerValAlaAla
3721  CCGGCTGCATATGCAGCTCAGGGCTATAAGGTGCTAGTACTCAACCCCTCTGTTGCTGCA
      GGCCGACGTATACGTCGAGTCCCGATATTCCACGATCATGAGTTGGGGAGACAACGACGT

      ThrLeuGlyPheGlyAlaTyrMetSerLysAlaHisGlyIleAspProAsnIleArgThr
3781  ACACTGGGCTTTGGTGCTTACATGTCCAAGGCTCATGGGATCGATCCTAACATCAGGACC
      TGTGACCCGAAACCACGAATGTACAGGTTCCGAGTACCCTAGCTAGGATTGTAGTCCTGG

      GlyValArgThrIleThrThrGlySerProIleThrTyrSerThrTyrGlyLysPheLeu
3841  GGGGTGAGAACAATTACCACTGGCAGCCCCATCACGTACTCCACCTACGGCAAGTTCCTT
      CCCCACTCTTGTTAATGGTGACCGTCGGGGTAGTGCATGAGGTGGATGCCGTTCAAGGAA

      AlaAspGlyGlyCysSerGlyGlyAlaTyrAspIleIleIleCysAspGluCysHisSer
3901  GCCGACGGCGGGTGCTCGGGGGGCGCTTATGACATAATAATTTGTGACGAGTGCCACTCC
```

Merriam-Webster's Collegiate® Dictionary

TENTH EDITION

Merriam-Webster, Incorporated
Springfield, Massachusetts, U.S.A.

A GENUINE MERRIAM-WEBSTER

The name *Webster* alone is no guarantee of excellence. It is used by a number of publishers and may serve mainly to mislead an unwary buyer.

Merriam-Webster™ is the name you should look for when you consider the purchase of dictionaries or other fine reference books. It carries the reputation of a company that has been publishing since 1831 and is your assurance of quality and authority.

Library of Congress Cataloging in Publication Data
Main entry under title:

Merriam-Webster's collegiate dictionary. — 10th ed.
p. cm.
Includes index.
ISBN 0-87779-708-0 (unindexed : alk. paper). — ISBN 0-87779-709-9 (indexed : alk. paper). — ISBN 0-87779-710-2 (deluxe indexed : alk. paper). — ISBN 0-87779-707-2 (laminated cover, unindexed).
1. English language—Dictionaries. I. Merriam-Webster, Inc.
PE1628.M36 1998
423—dc21 97-41846
CIP

Made in the United States of America

333435QWV02

pu·pil·age *or* **pu·pil·lage** \'pyü-pə-lij\ *n* (ca. 1599) : the state or period of being a pupil

pup·pet \'pə-pət\ *n, often attrib* [ME *popet*, fr. MF *poupette*, dim. of (assumed) *poupe* doll, fr. L *pupa*] (1538) **1 a** : a small-scale figure (as of a person or animal) usu. with a cloth body and hollow head that fits over and is moved by the hand **b** : MARIONETTE **2** : DOLL 1 **3** : one whose acts are controlled by an outside force or influence — **pup·pet-like** \-ˌlīk\ *adj*

pup·pe·teer \ˌpə-pə-'tir\ *n* (ca. 1923) : one who manipulates puppets

pup·pet·ry \'pə-pə-trē\ *n, pl* **-ries** (1528) **1** : the production or creation of puppets or puppet shows **2** : the art of manipulating puppets

pup·py \'pə-pē\ *n, pl* **puppies** [ME *popi*, fr. MF *poupée* doll, toy, fr. (assumed) *poupe* doll] (1591) : a young domestic dog; *specif* : one less than a year old — **pup·py·hood** \-ˌhůd\ *n* — **pup·py·ish** \-ish\ *adj* — **pup·py·like** \-ˌlīk\ *adj*

puppy dog *n* (1595) : a domestic dog; *esp* : one having the lovable attributes of a puppy

puppy love *n* (1834) : transitory affection felt by a boy or girl for one of the opposite sex

pup tent *n* (1863) : a low small tent for two persons usu. consisting of two halves fastened together

Pu·ra·na \pů-'rä-nə\ *n, often cap* [Skt *purāṇa*, fr. *purāṇa* ancient, fr. *purā* formerly; akin to Skt *pura* before, Gk *para* beside, *pro* before — more at FOR] (1696) : one of a class of Hindu sacred writings chiefly from A.D. 300 to A.D. 750 comprising popular myths and legends and other traditional lore — **Pu·ra·nic** \-nik\ *adj*

pur·blind \'pər-ˌblīnd\ *adj* [ME *pur blind*, fr. *pur* purely, wholly, fr. *pur* pure] (14c) **1 a** *obs* : wholly blind **b** : partly blind **2** : lacking in vision, insight, or understanding : OBTUSE — **pur·blind·ly** \-ˌblīn(d)-lē\ *adv* — **pur·blind·ness** \-ˌblīn(d)-nəs\ *n*

¹pur·chase \'pər-chəs\ *vb* **pur·chased; pur·chas·ing** [ME *purchacen*, fr. OF *purchacier* to seek to obtain, fr. *por-*, *pur-* for, forward (modif. of L *pro-*) + *chacier* to pursue, chase — more at PRO-] *vt* (14c) **1 a** *archaic* : GAIN, ACQUIRE **b** : to acquire (real estate) by means other than descent or inheritance **c** : to obtain by paying money or its equivalent : BUY **d** : to obtain by labor, danger, or sacrifice **2** : to apply a device for obtaining a mechanical advantage to (as something to be moved); *also* : to move by a purchase **3** : to constitute the means for buying ⟨our dollars ~ less each year⟩ ~ *vi* : to purchase something — **pur·chas·able** \-chə-sə-bəl\ *adj* — **pur·chas·er** *n*

²purchase *n* (14c) **1** : an act or instance of purchasing **2** : something obtained esp. for a price in money or its equivalent **3 a (1)** : a mechanical hold or advantage applied to the raising or moving of heavy bodies **(2)** : an apparatus or device by which advantage is gained **b (1)** : an advantage (as a firm hold or position) used in applying one's power ⟨clutching the steering wheel for more ~ —Barry Crump⟩ **(2)** : a means of exerting power

pur·dah \'pər-də\ *n* [Hindi *parda*, lit., screen, veil] (1865) **1** : seclusion of women from public observation among Muslims and some Hindus esp. in India **2** : a state of seclusion or concealment

pure \'pyůr\ *adj* **pur·er; pur·est** [ME *pur*, fr. OF, fr. L *purus;* akin to OHG *fowen* to sift, Skt *punāti* he cleanses, MIr *úr* fresh, new] (14c) **1 a (1)** : unmixed with any other matter ⟨~ gold⟩ **(2)** : free from dust, dirt, or taint ⟨~ food⟩ **(3)** : SPOTLESS, STAINLESS **b** : free from harshness or roughness and being in tune — used of a musical tone **c** *of a vowel* : characterized by no appreciable alteration of articulation during utterance **2 a** : being thus and no other : SHEER, UNMITIGATED ⟨~ folly⟩ **b (1)** : ABSTRACT, THEORETICAL **(2)** : A PRIORI ⟨~ mechanics⟩ **c** : not directed toward exposition of reality or solution of practical problems ⟨~ literature⟩ **d** : being nonobjective and to be appraised on formal and technical qualities only ⟨~ form⟩ **3 a (1)** : free from what vitiates, weakens, or pollutes **(2)** : containing nothing that does not properly belong **b** : free from moral fault or guilt **c** : marked by chastity : CONTINENT **d (1)** : of pure blood and unmixed ancestry **(2)** : homozygous in and breeding true for one or more characters **e** : ritually clean *syn* see CHASTE — **pure·ness** *n*

pure–blood·ed \'pyůr-ˌblə-dəd\ *or* **pure–blood** \-ˌbləd\ *adj* (1821) : FULL-BLOODED 1 — **pure·blood** \-ˌbləd\ *n*

pure·bred \-'bred, -ˌbred\ *adj* (1868) : bred from members of a recognized breed, strain, or kind without admixture of other blood over many generations — **pure·bred** \-ˌbred\ *n*

pure democracy *n* (ca. 1910) : democracy in which the power is exercised directly by the people rather than through representatives

¹pu·ree *or* **pu·rée** \pyů-'rā, -'rē\ *n* [F *purée*, fr. MF, fr. fem. of *puré*, pp. of *purer* to purify, strain, fr. L *purare* to purify, fr. *purus*] (1707) **1** : a paste or thick liquid suspension usu. made from cooked food ground finely **2** : a thick soup made of pureed vegetables

²puree *vt* **pu·reed; pu·ree·ing** (1928) : to make a puree of

pure imaginary *n* (1947) : a complex number that is the product of a real number other than zero and the imaginary unit — **pure imaginary** *adj*

pure·ly \'pyůr-lē\ *adv* (14c) **1** : WHOLLY, COMPLETELY ⟨a selection based ~ on merit⟩ **2** : without admixture of anything injurious or foreign **3** : SIMPLY, MERELY ⟨read ~ for relaxation⟩ **4** : in a chaste or innocent manner

pur·fle \'pər-fəl\ *vt* **pur·fled; pur·fling** \-f(ə-)liŋ\ [ME *purfilen*, fr. MF *porfiler*, fr. (assumed) VL *profilare*, fr. L *pro-* forward + LL *filare* to spin — more at PRO-, FILE] (14c) : to ornament the border or edges of — **purfle** *n*

pur·ga·tion \ˌpər-'gā-shən\ *n* (14c) : the act or result of purging

¹pur·ga·tive \'pər-gə-tiv\ *adj* [ME *purgatif*, fr. MF, fr. LL *purgativus*, fr. L *purgatus*, pp.] (15c) : purging or tending to purge

²purgative *n* (1626) : a purging medicine : CATHARTIC

pur·ga·to·ri·al \ˌpər-gə-'tōr-ē-əl, -'tȯr-\ *adj* (15c) **1** : of, relating to, or suggestive of purgatory **2** : cleansing of sin : EXPIATORY

pur·ga·to·ry \'pər-gə-ˌtōr-ē, -ˌtȯr-\ *n, pl* **-ries** [ME, fr. AF or ML; AF *purgatorie*, fr. ML *purgatorium*, fr. LL, neut. of *purgatorius* purging, fr. L *purgare*] (13c) **1** : an intermediate state after death for expiatory purification; *specif* : a place or state of punishment wherein according to Roman Catholic doctrine the souls of those who die in God's grace may make satisfaction for past sins and so become fit for heaven **2** : a place or state of temporary suffering or misery

¹purge \'pərj\ *vb* **purged; purg·ing** [ME, fr. MF *purgier*, fr. L *purigare*, *purgare* to purify, purge, fr. *purus* pure + *-igare* (akin to *agere* to drive, do) — more at ACT] *vt* (14c) **1 a** : to clear of guilt **b** : to free from moral or ceremonial defilement **2 a** : to cause evacuation from (as the bowels) **b (1)** : to make free of something unwanted ⟨~ a manhole of gas⟩ ⟨~ yourself of fear⟩ **(2)** : to free (as a boiler) of sediment or relieve (as a steam pipe) of trapped air by bleeding **c (1)** : to rid (as a nation or party) by a purge **(2)** : to get rid of ⟨the leaders had been *purged*⟩ ⟨~ money-losing operations⟩ ~ *vi* **1** : to become purged **2** : to have or produce frequent evacuations **3** : to cause purgation — **purg·er** *n*

²purge *n* (1563) **1** : something that purges; *esp* : PURGATIVE **2 a** : an act or instance of purging **b** : the removal of elements or members regarded as undesirable and esp. as treacherous or disloyal

pu·ri \'půr-ē\ *n, pl* **puri** *or* **puris** [Hindi *pūrī*, fr. Skt *pūra*] (1839) : a puffy fried wheat cake of India

pu·ri·fi·ca·tion \ˌpyůr-ə-fə-'kā-shən\ *n* (14c) : the act or an instance of purifying or of being purified

pu·ri·fi·ca·tor \'pyůr-ə-fə-ˌkā-tər\ *n* (1853) **1** : a linen cloth used to wipe the chalice after celebration of the Eucharist **2** : one that purifies

pu·ri·fi·ca·to·ry \pyůr-'i-fi-kə-ˌtōr-ē, 'pyůr-(ə-)fə-kə-, -ˌtȯr-\ *adj* (1610) : serving, tending, or intended to purify

pu·ri·fy \'pyůr-ə-ˌfī\ *vb* **-fied; -fy·ing** [ME *purifien*, fr. MF *purifier*, fr. L *purificare*, fr. L *purus* + *-ificare* -ify] *vt* (14c) : to make pure: as **a** : to clear from material defilement or imperfection **b** : to free from guilt or moral or ceremonial blemish **c** : to free from undesirable elements ~ *vi* : to grow or become pure or clean — **pu·ri·fi·er** \-ˌfī(-ə)r\ *n*

Pu·rim \'půr-im, 'pyůr-, -ˌēm; pů-'rim, pyů-, -'rēm\ *n* [Heb *pūrīm*, lit., lots; fr. the casting of lots by Haman (Esth 9:24–26)] (1535) : a Jewish holiday celebrated on the 14th of Adar in commemoration of the deliverance of the Jews from the massacre plotted by Haman

pu·rine \'pyůr-ˌēn\ *n* [G *Purin*, fr. L *purus* pure + NL *uricus* uric (fr. E *uric*) + G *-in* ²-ine] (1898) **1** : a crystalline base $C_5H_4N_4$ that is the parent of compounds of the uric-acid group **2** : a derivative of purine; *esp* : a base (as adenine or guanine) that is a constituent of DNA or RNA

pur·ism \'pyůr-ˌi-zəm\ *n* (1803) **1** : an example of rigid adherence to or insistence on purity or nicety esp. in use of words; *esp* : a word, phrase, or sense used chiefly by purists **2** : the quality or practice of adherence to purity esp. in language

pur·ist \'pyůr-ist\ *n* (ca. 1706) : one who adheres strictly and often excessively to a tradition; *esp* : one preoccupied with the purity of a language and its protection from the use of foreign or altered forms — **pu·ris·tic** \pyů-'ris-tik\ *adj* — **pu·ris·ti·cal·ly** \-ti-k(ə-)lē\ *adv*

¹pu·ri·tan \'pyůr-ə-tᵊn\ *n* [prob. fr. LL *puritas* purity] (1572) **1** *cap* : a member of a 16th and 17th century Protestant group in England and New England opposing as unscriptural the ceremonial worship and the prelacy of the Church of England **2** : one who practices or preaches a more rigorous or professedly purer moral code than that which prevails

²puritan *adj, often cap* (1589) : of or relating to puritans, the Puritans, or puritanism

pu·ri·tan·i·cal \ˌpyůr-ə-'ta-ni-kəl\ *adj* (1607) **1** : PURITAN **2** : of, relating to, or characterized by a rigid morality — **pu·ri·tan·i·cal·ly** \-k(ə-)lē\ *adv*

pu·ri·tan·ism \'pyůr-ə-tᵊn-ˌi-zəm\ *n* (1573) **1** *cap* : the beliefs and practices characteristic of the Puritans **2** : strictness and austerity esp. in matters of religion or conduct

pu·ri·ty \'pyůr-ə-tē\ *n* [ME *purete*, fr. OF *pureté*, fr. LL *puritat-*, *puritas*, fr. L *purus* pure] (13c) **1** : the quality or state of being pure **2** : SATURATION 4a

Pur·kin·je cell \(ˌ)pər-'kin-jē-\ *n* [Jan *Purkinje*] (ca. 1890) : any of numerous nerve cells that occupy the middle layer of the cerebellar cortex and are characterized by a large globose body with massive dendrites directed outward and a single slender axon directed inward

Purkinje fiber *n* (ca. 1890) : any of the modified cardiac muscle fibers that have few nuclei, granulated central cytoplasm, and sparse peripheral striations and make up a network of conducting tissue in the myocardium

¹purl \'pər(-ə)l\ *n* [ME] (14c) **1** : gold or silver thread or wire for embroidering or edging **2** : the intertwisting of thread that knots a stitch usu. along an edge **3** : PURL STITCH

²purl *vt* (1526) **1 a** : to embroider with gold or silver thread **b** : to edge or border with gold or silver embroidery **2** : to knit in purl stitch ~ *vi* : to do knitting in purl stitch

³purl *n* [perh. of Scand origin; akin to Norw *purla* to ripple] (ca. 1522) **1** : a purling or swirling stream or rill **2** : a gentle murmur or movement (as of purling water)

⁴purl *vi* (1591) **1** : EDDY, SWIRL **2** : to make a soft murmuring sound like that of a purling stream

pur·lieu \'pərl-(ˌ)yü, 'pər-(ˌ)lü\ *n* [ME *purlewe* land severed from an English royal forest by perambulation, fr. AF *puralé* perambulation, fr. OF *puraler* to go through, fr. *pur-* for, through + *aler* to go — more at PURCHASE] (15c) **1 a** : an outlying or adjacent district **b** *pl* : ENVIRONS, NEIGHBORHOOD **2 a** : a frequently visited place : HAUNT **b** *pl* : CONFINES, BOUNDS

pur·lin \'pər-lən\ *n* [origin unknown] (15c) : a horizontal member in a roof

pur·loin \(ˌ)pər-'lȯin, 'pər-ˌ\ *vt* [ME, to put away, misappropriate, fr. AF *purloigner*, fr. OF *porloigner* to put off, delay, fr. *por-* forward + *loing* at a distance, fr. L *longe*, fr. *longus* long — more at PURCHASE, LONG] (15c) : to appropriate wrongfully and often by a breach of trust *syn* see STEAL — **pur·loin·er** *n*

purl stitch *n* [¹*purl*] (1885) : a knitting stitch usu. made with the yarn at the front of the work by inserting the right needle into the front of a loop on the left needle from the right, catching the yarn with the right needle, and bringing it through to form a new loop — compare KNIT STITCH

pu·ro·my·cin \ˌpyůr-ə-'mī-sᵊn\ *n* [*purine* + *-o-* + *-mycin*] (1953) : an antibiotic $C_{22}H_{29}N_7O_5$ that is obtained from an actinomycete (*Streptomyces alboniger*) and is used esp. as a potent inhibitor of protein synthesis

¹pur·ple \'pər-pəl\ *adj* **pur·pler** \-p(ə-)lər\; **pur·plest** \-p(ə-)ləst\ [ME *purpel*, alter. of *purper*, fr. OE *purpuran* of purple, gen. of *purpure* purple color, fr. L *purpura*, fr. Gk *porphyra*] (bef. 12c) **1** : REGAL, IMPERIAL **2** : of the color purple **3 a** : highly rhetorical : ORNATE **b** : marked by profanity

NANBV DIAGNOSTICS AND VACCINES

Cross-Reference to Related Applications

This application is a continuation-in-part of applicants' copending applications: attorney docket number 2300-0063 (formerly 2300-0203, U.S.S.N. 122,714), filed 18 November 1987; attorney docket number 2300-0063.20 (formerly 2300-0219, U.S.S.N. 139,886), filed 30 December 1987; attorney docket number 2300-0063.21 (formerly 2300-0228, U.S.S.N. 161,072), filed 26 February 1988; attorney docket number 2300-0063.22 (formerly 2300-0237, U.S.S.N. 191,263), filed 6 May 1988; attorney docket number 2300-0063.23 (U.S.S.N. 263,584, filed 26 October 1988); attorney docket number 2300-0063.24 (U.S.S.N. 271,450), filed 14 November 1988; attorney docket number 2300-0063.59 (PCT/US88/04125, filed 18 November 1988, converted to U.S. National phase on April 21, 1989 and assigned attorney docket number 2300-0063.26); attorney docket number 2300-0063.59 (U.S.S.N. 07/325,338, filed 17 March 1989); and attorney docket number 2300-0063.29 (U.S.S.N. 07/341,334 filed 20 April 1989), which are incorporated herein by reference.

Technical Field

The invention relates to materials and methodologies for managing the spread of non-A, non-B hepatitis virus (NANBV) infection. More specifically, it relates to diagnostic DNA fragments, diagnostic proteins, diagnostic antibodies and protective antigens and antibod-

Fig. 52 shows the nucleotide sequence of HCV cDNA in clone CA156e, the amino acids encoded therein, and the sequences which overlap with CA84a.

Fig. 53 shows the nucleotide sequence of HCV cDNA in clone CA167b, the amino acids encoded therein, and the sequences which overlap CA156e.

Fig. 54 shows the ORF of HCV cDNA derived from clones pi14a, CA167b, CA156e, CA84a, CA59a, K9-1, 12f, 14i, 11b, 7f, 7e, 8h, 33c, 40b, 37b, 35, 36, 81, 32, 33b, 25c, 14c, 8f, 33f, 33g, 39c, 35f, 19g, 26g, and 15e.

Fig. 55 shows the hydrophobicity profiles of polyproteins encoded in HCV and in West Nile virus.

Fig. 56 shows the nucleotide sequence of HCV cDNA in clone CA216a, the amino acids encoded therein, and the overlap with clone CA167b.

Fig. 57 shows the nucleotide sequence of HCV cDNA in clone CA290a, the amino acids encoded therein, and the overlap with clone CA216a.

Fig. 58 shows the nucleotide sequence of HCV cDNA in clone ag30a and the overlap with clone CA290a.

Fig. 59 shows the nucleotide sequence of HCV cDNA in clone CA205a, and the overlap with the HCV cDNA sequence in clone CA290a.

Fig. 60 shows the nucleotide sequence of HCV cDNA in clone 18g, and the overlap with the HCV cDNA sequence in clone ag30a.

Fig. 61 shows the nucleotide sequence of HCV cDNA in clone 16jh, the amino acids encoded therein, and the overlap of nucleotides with the HCV cDNA sequence in clone 15e.

Fig. 62 shows the composite sequence of the HCV cDNA sense strand deduced from overlapping clones b114a, 18g, ag30a, CA205a, CA290a, CA216a, pi14a, CA167b, CA156e, CA84a, CA59a, K9-1 (also called k9-1), 26j, 13i, 12f, 14i, 11b, 7f, 7e, 8h, 33c, 40b, 37b, 35, 36, 81, 32, 33b, 25c, 14c, 8f, 33f, 33g, 39c, 35f, 19g, 26g, 15e, b5a, and 16jh.

individual NS proteins in the putative Flavirus precursor polyprotein are fairly well-known. Moreover, these also coincide with observed gross fluctuations in the hydrophobicity profile of the polyprotein. It is established that NS5 of Flaviviruses encodes the virion polymerase, and that NS1 corresponds with a complement fixation antigen which has been shown to be an effective vaccine in animals. Recently, it has been shown that a flaviviral protease function resides in NS3. Due to the observed similarities betwen HCV and the Flaviviruses, deductions concerning the approximate locations of the corresponding protein domains and functions in the HCV polyprotein are possible (See Section IV.H.6). The expression of polypeptides containing these domains in a variety of recombinant host cells, including, for example, bacteria, yeast, insect, and vertebrate cells, should give rise to important immunological reagents which can be used for diagnosis, detection, and vaccines.

Although the non-structural protein region of the putative polyproteins of the HCV isolate described herein and of Flaviviruses appears to be generally similar, there is less similarity between the putative structural regions which are towards the N-terminus. In this region, there is a greater divergence in sequence, and in addition, the hydrophobic profile of the two regions show less similarity. This "divergence" begins in the N-terminal region of the putative NS1 domain in HCV, and extends to the presumed N-terminus. Nevertheless, it is still possible to predict the approximate locations of the putative nucleocapsid (N-terminal basic domain) and E (generally hydrophobic) domains within the HCV polyprotein. In Section IV.H.6., the predictions are based on the changes observed in the hydrophobic profile of the HCV polyprotein, and on a knowledge of the location and character of the flaviviral proteins. From these predictions it may be possible to

upon the Flavivirus model and the hydropathic character of the putative encoded polypeptides. However, the hydrophobicity profiles (described infra), indicate that HCV diverges from the Flavivirus model, particularly with respect to the region upstream of NS2. Moreover, the boundaries indicated are not intended to show firm demarcations between the putative polypeptides.

The possible protein domains of the encoded HCV polyprotein, as well as the approximate boundaries, are the following:

Putative Domain	Approximate Boundary (amino acid nos.)
C (nucleocapsid protein)	1-120
E (Virion envelope protein(s) and possibly matrix (M) proteins	120-400
NS1 (complement fixation antigen?)	400-660
NS2 (unknown function)	660-1050
NS3 (protease?)	1050-1640
NS4 (unknown function)	1640-2000
NS5 (polymerase)	2000-? end

The expression vectors containing the cloned HCV cDNAs were constructed from pSODcf1, which is described in Section IV.B.1. In order to be certain that a correct reading frame would be achieved, three separate expression

SEQUENCE OF THE HCV cDNA SENSE STRAND

(AS DEDUCED FROM OVERLAPPING CLONES b114a/18g/ag30a/CA203a/CA290a/CA216a/pi14a/CA167b/CA156e/CA84a/CA59a/K9-1/26j/13i/12f/141/11b/7f/7e/8h/33c/40b/37b/35/36/81/32/33b/25c/14c/8f/33f/33g/39c/35f/19g/26g/15e/b5a/16jh)

CACTCCACCATGAATCACTCCCCTGTGAGGAACTACTGTCTTCACGCAGAAAGCGTCTAG
CCATGGCGTTAGTATGAGTGTCGTGCAGCCTCCAGGACCCCCCCTCCCGGGAGAGCCATA
GTGGTCTGCGGAACCGGTGAGTACACCGGAATTGCCAGGACGACCGGGTCCTTTCTTGGA
TCAACCCGCTCAATGCCTGGAGATTTGGGCGTGCCCCCGCAAGACTGCTAGCCGAGTAGT
GTTGGGTCGCGAAAGGCCTTGTGGTACTGCCTGATAGGGTGCTTGCGAGTGCCCCGGGAG-300

---(Putative initiator methionine codon)

GTCTCGTAGACCGTGCACCATGAGCACGAATCCTAAACCTCAAAAAAAAAAACAAACGTAA
CACCAACCGTGGCCCACAGGACGTCAAGTTCCCGGGTGGCGGTCAGATCGTTGGTGGAGT
TTACTTGTTGCCCGCGCAGGGGCCCTAGATTGGGTGTGCGCGCGACGAGAAAGACTTCCGA
GCGGTCGCAACCTCGAGGTAGACGTCAGCCTATCCCCAAGGCTCGTCGGCCCCGAGGGCAC
GACCTCCGCTCAGCCCGCGTACCCTTGGCCCCTCTATGGCAATGAGGGCTGCGGGTGGGC-600
GGGATGGCTCCTGTCTCCCCGTGGCTCTCGGCCTAGCTGGGGCCCCACAGACCCCCGCCC
TACGTCGCGCAATTTCCGTAAGGTCATCGATACCCTTACGTGCGGCTTCGCCGACCTCAT
GGGGTACATACCGCTCGTCGGCGCCCCTCTTGGAGGCGCTGCCAGGGCCCTGGCGCATGG
CGTCCGGGTTCTGGAAGACGGCGTGAACTATCCAACAGGCAACCTTCCTGGTTGCTCTTT
CTCTATCTTCCTTCTGGCCCTGCTCTCTTGCTTGACTGTGCCCGCTTCGGCCTACCAAGT-900
GCGCAACTCCACGGGGCTTTACCACGTCACCAATGATTGCCCTAACTCCAGTATTGTGTA
CGAGGCGGCCGATCCCATCCTGCACACTCCGGGGTGCGTCCCTTGCGTTCGTGAGGGCAA
CGCCTCGAGGTGTTGGGTGGCGATGACCCCTACGGTGGCCACCAGGGATGCCAAACTCCC
CGCGACGCACCTTCGACGTCACATCGATCTGCTTGTCGGGAGCGCCACCCTCTGTTCGGC
CCTCTACGTGGGGGACCTATGCGGGTCTGTCTTTCTTGTCGGCCAACTGTTCACCTTCTC-1200
TCCCAGGCGCCACTGGACGACCCAAGGTTGCAATTGCTCTATCTATCCCGGCCATATAAC
GGGTCACCGCATGGCATGGGATATGATGATGAACTGGTCCCCTACGACGGCGTTGGTAAT
GGCTCAGCTGCTCCGGATCCCACAAGCCATCTTCGACATCATCGCTGGTGCTCACTGGGG
AGTCCTGGCGGGCATAGCGTATTTCTCCATGGTGGGGAACTGGGCGAAGGTCCTGGTAGT
GCTGCTGCTATTTGCCGGCGTCGACGCGGAAACCCACGTCACCGGCGCAAGTGCCGCCCA-1500
CACTGTGTCTGCATTTGTTAGCCTCCTCGCACCAGGCGCCAAGCAGAACGTCCAGCTGAT
CAACACCAACGGCAGTTGGCACCTCAATAGCACGGCCCTGAACTGCAATGATAGCCTCAA
CACCGCCTGGTTGGCAGGCTTTTCTATCACCACAAGTTCAACTCTTCAGGCTGTCCTGA
GAGGCTAGCCAGCTGCCGACCCCTTACCGATTTTGACCAGGGCTGGGGCCCTATCAGTTA
TGCCAACGGAAGCGGCCCCGACCAGCGCCCCTACTGCTGGCACTACCCCCCAAAACCTTG-1800
CGGTATTGTGCCCGCGAAGAGTGTGTGTGGTCCGGTATATTGCTTCACTCCCAGCCCCGT
GGTGGTGGGAACGACCGACAGGTCGGGCGCGCCCACCTACAGCTGGCGTCAAAATGATAC
GGACGTCTTCGTCCTTAACAATACCAGGCCACCGCTGGGCAATTGGTTCGGTTGTACCTG
GATGAACTCAACTGGATTCACCAAAGTGTGCGGAGCGCCTCCTTGTGTCATCGGAGGCGC
GGGCAACAACACCCTGCACTGCCCCACTGATTGCTTCCGCAAGCATCCGGACGCCACATA-2100
CTCTCGGTGCGGCTCCGGTCCCTGGATCACACCCAGGTGCCTGGTCGACTACCCGTATAG
GCTTTGCCATTATCCTTGTACCATCAACTACACCATATTTAAAATCAGGATGTACGTGGG
AGGGGTCGAACACAGGCTGGAAGCTGCCTGCAACTGGACGCGGGGCGAACGTTGCGATCT
GGAAGACAGGGACAGGTCCGAGCTCAGCCCGTTACTGCTGACCACTACACACTGGCAGGT
CCTCCCGTGTTCCTTCACAACCCTACCAGCCTTGTCCACCGGCCTCATCCACCTCCACCA-2400
GAACATTGTGGACGTGCAGTACTTGTACGGGGTGGGGTCAAGCATCGCGTCCTGGGCCAT
TAACTGCGAGTACGTCGTTCTCCTGTTCCTTCTGCTTGCAGACGCGCGCGTCTGCTCCTG
CTTGTGGATGATGCTACTCATATCCCAAGCGGAGGCGGCTTTGGAGAACCTCGTAATACT
TAATGCAGCATCCCTCGCCCGGGACGCACGGTCTTGTATCCTTCCTCGTGTTCTTCTGCTT
TGCATGGTATTTGAAGGGTAAGTGGGTGCCCGGAGCGGTCTACACCTTCTACGGGATGTG-2700
GCCTCTCCTCCTGCTCCTGTTGGCGTTGCCCCACCGGCCGTACGCGCTGGACACCGACGT
GGCCGCGTCGTGTGGCGGTGTTGTTCTCGTCGGGTTGATGGCGCTGACTCTGTCACCATA
TTACAAGCGCTATATCAGCTGGTGCTTGTGGTGGCTTCAGTATTTCTGACCAGACTGGA
ACCGCAACTGCACCTGTGCATTCCCCCCCTCAACGTCCGAGGGGGGCGCGACGCCGTCAT
CTTACTCATGTGTGCTGTACACCCGACTCTGGTATTTGACATCACCAAATTGCTGCTGGC-3000
CGTCTTCGGACCCCTTTGGATTCTTCAAGCCAGTTTGCTTAAAGTACCCTACTTTGTGCG
CGTCCAAGGCCTTCTCCGGTTCTGCGCGTTAGCGCGGAAGATGATCGGAGGCCATTACGT
GCAAATGGTCATCATTAAGTTACGGGCGCTTACTGGCACCTATGTTTATAACCATCTCAC
TCCTCTTCGGGACTGGGCGCACAACGGCTTGCGAGATCTGGCCGTGGCTGTAGAGCCAGT
CGTCTTCTCCCAAATGGAGACCAAGCTCATCACGTGGGGGGCCAGATACCGCCGCGTGCGG-3300

FIGURE 62

TGACATCATCAACGGCTTGCCTGTTTCCGCCCCGCAGGGGCCGGGAGATACTGCTCGGGCC
ACCCCATGGAATGGTCTCCAAGGGGTGGAGGTTGCTGGCGCCCATCACGGCGTACGCCCA
GCAGACAAGGGGCCTCCTAGGGTGCATAATCACCAGCCTAACTGGCCGGGACAAAAACCA
AGTCCACGCTCACGTCCAGATTGTGTCAACTGCTGCCCAAACCTTCCTGGCAACGTGCAT
CAATGGGGTGTGCTGGACTGTCTACCACGGGGCCCCAACGACGACCATCGCGTCACCCAA-3600
GGGTCCTGTCATCCACATCTATACCAATGTAGACCAAGACCTTGTGGGCTGGCCCGCTCC
GCAAGGTAGCCGCTCATTGACACCCTGCACTTGCGGCTCCTCGGACCTTTACCTGGTCAC
GAGGCACCCCGATCTCATTCCCGTGCGCCGGCGGGGTGATAGCAGGGGCAGCCTGCTGTC
GCCCCGGCCCATTTCCTACTTGAAAGGCTCCTCGGGGGGTCCGCTCTTGTGCCCCGCGGC
GCACGCCGTGGGCATATTTAGGCCCCGGGTCTGCACCCCGTGGAGTGGCTAAGGCGGTGGA-3900
CTTTATCCCTGTGGAGAACCTAGAGACAACCATGAGGTCCCCGGTGTTCACCGATAACTC
CTCTCCACCAGTAGTGCCCCAGAGCTTCCAGGTGCCTCACCTCCATGCTCCCACAGGCAG
CGGCAAAAGCACCAAGGTCCCGGCTGCATATGCAGCTCAGGGCTATAAGGTGCTAGTACT
CAACCCCTCTGTTGCTGCAACACTGGCCTTTGGTCCTTACATGTCCAAGGCTCATGGGAT
CGATCCTAACATCAGGACCGGGGTGAGAACAATTACCACTGGCAGCCCCATCACGTACTC-4200
CACCTACGGCAAGTTCCTTGCCGACGGCGGGTCCTCGGGGCCCGCTTATGACATAATAAT
TTGTGACCAGTGCCACTCCACGGATGCCACATCCATCTTGGGCATCGGCACTGTCCTTGA
CCAAGCAGAGACTGCGGGGGCGAGACTGGTTGTGCTCGCCACCGCCACCCCTCCGCGCTC
CGTCACTGTGCCCCATCCCAACATCGAGGAGGTTGCTCTGTCCACCACCGGAGAGATCCC
TTTTTACGGCAAGGCTATCCCCCTCGAAGTAATCAAGGGGGGGAGACATCTCATCTTCTC-4500
TCATTCAAACAAGAAGTCCGACGAACTCGCCGCAAAGCTGGTCGCATTGGGCATCAATGC
CGTGGCCTACTACCGCGGTCTTGACGTGTCCGTCATCCCGACCAGCGGCGATGTTGTCGT
CGTGGCAACCGATGCCCTCATGACCGGCTATACCGCCGACTTCGACTCGGTGATAGACTG
CAATACGTGTGTCACCCAGACAGTCGATTTCAGCCTTGACCCTACCTTCACCATTGAGAC
AATCACGCTCCCCCAGGATGCTGTCTCCCGCACTCAACGTCGCGGCAGCACTGGCAGGGG-4800
GAAGCCAGGCATCTACAGATTTGTGGCACCGGGGGAGCGCCCCTCCGGCATGTTCGACTC
GTCCGTCCTCTGTGAGTGCTATGACGCAGGCTGTGCTTGGTATGAGCTCACGCCCCGCCGA
GACTACAGTTAGGCTACGAGCGTACATGAACACCCCGGGGCTTCCCGTGTGCCAGGACCA
TCTTGAATTTTGGGAGGGCGTCTTTACAGGCCTCACTCATATAGATGCCCACTTTCTATC
CCAGACAAAGCAGAGTGGGGAGAACCTTCCTTACCTGGTAGCGTACCAAGCCACCGTGTG-5100
CGCTAGGGCTCAAGCCCCTCCCCCATCGTGGGACCAGATGTGGAAGTGTTTGATTCGCCT
CAAGCCCACCCTCCATGGGCCAACACCCCTGCTATACAGACTGGGCGCTGTTCAGAATGA
AATCACCCTGACCCACCCAGTCACCAAATACATCATGACATGCATGTCGGCCGACCTGGA
GGTCGTCACGAGCACCTGGGTGCTCGTTGGCGGCGTCCTGGCTGCTTTGGCCGCGTATTG
CCTGTCAACAGGCTGCGTGCTCATAGTGCGCCAGCCTCGTCTTGTCCGGGAAGCCGGCAAT-5400
CATACCTGACAGGGAAGTCCTCTACCGAGAGTTCGATGAGATGGAAGAGTGCTCTCAGCA
CTTACCGTACATCGAGCAAGGGATGATGCTCGCCCAGCAGTTCAAGCAGAAGGCCCTCGG
CCTCCTGCAGACCGCGTCCCGTCAGGCAGAGGTTATCGCCCCTGCTGTCCAGACCAACTG
GCAAAAACTCGAGACCTTCTGGGCGAAGCATATGTGGAACTTCATCAGTGGGATACAATA
CTTGGCGGCCTTGTCAACGCTGCCTGGTAACCCCGCCATTGCTTCATTGATGGCTTTTAC-5700
AGCTGCTGTCACCAGCCCACTAACCACTAGCCAAACCCTCCTCTTCAACATATTGGGGGG
GTGGGTGGCTGCCCAGCTCGCCGCCCCCGGTGCCGCTACTGCCTTTGTGGGCGCTGGCTT
AGCTGGCGCCGCCATCGGCAGTGTTGGACTGGGGAAGGTCCTCATAGACATCCTTGCAGG
GTATGGCGCGGGCGTGGCGGGAGCTCTTGTGGCATTCAAGATCATCAGCGGTGAGGTCCC
CTCCACGGAGGACCTGGTCAATCTACTGCCCGCCATCCTCTCGCCCGGAGCCCTCGTAGT-6000
CGGCGTGGTCTGTGCAGCAATACTGCGCCGGCACGTTGGCCCCGGCGAGCGGGCAGTGCA
GTCCATGAACCGGCTGATAGCCTTCGCCTCCCGGGGGAACCATGTTCCCCCACGCACTA
CGTGCCGGAGAGCGATGCAGCTGCCCGCGTCACTGCCATACTCAGCAGCCTCACTGTAAC
CCAGCTCCTGAGGCGACTCCACCAGTGGATAAGCTCGGAGTGTACCACTCCATGCTCCGG
TTCCTGGCTAAGGGACATCTGGGACTGGATATGCGAGGTGTTGAGCGACTTTAAGACCTG-6300
GCTAAAAGCTAAGCTCATGCCACAGCTGCCTGGGATCCCCTTTGTGTCCTGCCAGCGCGG
GTATAAGGGGGTCTGGCGAGTGGACGGCATCATGCACACTCGCTGCCACTGTGGAGCTGA
GATCACTGGACATGTCAAAAACGGGACGATGAGGATCGTCGGTCCTAGGACCTGCAGCAA
CATGTGGAGTGGGACCTTCCCCATTAATGCCTACACCACGGGCCCCTGTACCCCCCTTCC
TGCGCCGAACTACACGTTCGCGCTATGGAGGGTGTCTGCAGACGAATATGTGGAGATAAG-6600
GCAGCTGGGGGACTTCCACTACGTGACGGGTATGACTACTGACAATCTCAAATGCCCGTG
CCAGGTCCCATCGCCCGAATTTTTCACAGAATTGGACGGGGTGCGCCTACATAGGTTTGC
GCCCCCCTGCAAGCCCTTGCTGCGCGAGGAGGTATCATTCAGAGTAGGACTCCACGAATA
CCCGGTAGGGTCGCAATTACCTTGCGAGCCCGAACCGGACGTGGCCGTGTTGACGTCCAT
CCTCACTGATCCCTCCCATATAACAGCAGAGGCGGCCGGCCGAAGGTTGGCGAGGGGATC-6900
ACCCCCCTCTGTGGCCAGCTCCTCGGCTAGCCAGCTATCCGCTCCATCTCTCAAGGCAAC
TTGCACCGCTAACCATGACTCCCCTGATGCTGAGCTCATAGAGGCCAACCTCCTATGGAG
GCAGGAGATGGGCGGCAACATCACCAGGGTTGAGTCAGAAAACAAAGTGGTGATTCTGGA
CTCCTTCGATCCGCTTGTGGCGGAGGAGGACGAGCGGGAGATCTCCGTACCCGCAGAAAT
CCTGCGGAAGTCTCGGAGATTCGCCCAGGCCCTGCCCGTTTGGGCGCGGCCGGACTATAA-7200
CCCCCCGCTAGTGGAGACGTGGAAAAAGCCCGACTACGAACCACCTGTGGTCCATGGCTG

FIGURE 62.1

TCCGCTTCCACCTCCAAAGTCCCCTCCTGTGCCTCCGCCTCGGAAGAAGCGGACGGTGGT
CCTCACTGAATCAACCCTATCTACTGCCTTGGCCGAGCTCGCCACCAGAAGCTTTGGCAG
CTCCTCAACTTCCGGCATTACGGGCGACAATACGACAACATCCTCTCAGCCCGCCCCTTC
TGCCTGCCCCCCCGACTCCGACGCTGAGTCCTATTCCTCCATGCCCCCCCTGGAGGGGGA-7500
GCCTGGGGATCCGGATCTTAGCGACGGGTCATGGTCAACGGTCAGTAGTCAGGCCAACGC
GGAGGATGTCGTGTCCTGCTCAATGTCTTACTCTTGGACAGGCGCACTCGTCACCCCGTG
CGCCGCGGAAGAACAGAAACTGCCCATCAATGCACTAAGCAACTCGTTGCTACGTCACCA
CAATTTGGTGTATTCCACCACCTCACGCAGTGCTTGCCAAAGGCAGAAGAAAGTCACATT
TGACAGACTGCAAGTTCTGGACAGCCATTACCAGGACGTACTCAAGGAGGTTAAAGCAGC-7800
GGCGTCAAAAGTGAAGGCTAACTTCCTATCCGTAGAGGAAGCTTGCAGCCTGACGCCCCC
ACACTCAGCCAAATCCAAGTTTGGTTATGGGGCAAAAGACGTCCGTTGCCATGCCAGAAA
GGCCGTAACCCACATCAACTCCGTCTGGAAACACCTTCTCGAAGACAATGTAACACCAAT
AGACACTACCATCATGGCTAAGAACGAGGTTTTCTGCGTTCAGCCTGAGAAGGGGGGTCG
TAAGCCAGCTCGTCTCATCGTGTTCCCCGATCTGGGCCTGCGCCGTGTGCGAAAAGATGGC-8100
TTTCTACGACCTGGTTACAAAGCTCCCCTTGGCCGTGATGGGAAGCTCCTACGGATTCCA
ATACTCACCAGGACAGCGGGTTGAATTCCTCGTGCAAGCGTGGAAGTCCAAGAAAACCCC
AATGGCGTTCTCGTATGATACCCCCTGCTTTGACTCCACAGTCACTGAGAGCGACATCCG
TACGGAGGAGGCAATCTACCAATGTTGTGACCTCGACCCCCAAGCCCGCGTGGCCATCAA
GTCCCTCACCGAGAGGCTTTATGTTGGGGGCCCTCTTACCAATTCAAGGGGGGAGAACTG-8400
CGGCTATCGCAGGTGCCGCGCGAGCGGCGTACTGACAACTAGCTGTGGTAACACCCTCAC
TTGCTACATCAAGGCCCGGGCACCCTGTCGACCCGCACGGCTCCAGGACTGCACCATGCT
CGTGTGTGGCGACGACTTAGTCGTTATCTGTGAAAGCGCGGGGGTCCAGGAGGACGCGGC
GAGCCTGAGAGCCTTCACGGAGGCTATGACCAGGTACTCCGCCCCCCCTGGGGACCCCCC
ACAACCAGAATACGACTTGCAGCTCATAACATCATGCTCCTCCAACGTGTCAGTCGCCCA-8700
CGACGGCGCTGGAAAGAGGGTCTACTACCTCACCCGTGACCCTACAACCCCCCTCGCGAG
AGCTGCGTGGGAGACAGCAAGACACACTCCAGTCAATTCCTGGCTAGGCAACATAATCAT
GTTTCCCCCCACACTGTGGGCGAGGATGATACTGATGACCCATTTCTTTAGCGTCCTTAT
AGCCAGGGACCAGCTTGAACAGGCCCTCGATTGCGAGATCTACGCGCCCTGCTACTCCAT
AGAACCACTTGATCTACCTCCAATCATTCAAAGACTCCATGGCCTCAGCGCATTTTCACT-9000
CCACAGTTACTCTCCAGGTGAAATTAATAGGGTGGCCGCATGCCTCAGAAACTTGCGGGT
ACCGCCCTTGCGAGCTTGGAGACACCGCGCCCCGGACCGTCCGCGCTAGGCTTCTGGCCAG
AGGAGGCAGGGCTGCCATATGTGGCAAGTACCTCTTCAACTGGGCAGTAAGAACAAAGCT
CAAAC

FIGURE 62.2

COMBINED DECLARATION AND POWER OF ATTORNEY FOR CONTINUATION-IN-PART APPLICATION	Attorney Docket No. 2300-0063.28

As a below named inventor, I hereby declare that:
My residence, post office address and citizenship are as stated below next to my name;
I BELIEVE I AM THE ORIGINAL, FIRST AND SOLE INVENTOR (if only one name is listed below) OR AN ORIGINAL, FIRST AND JOINT INVENTOR (if more than one name is listed below) OF THE SUBJECT MATTER WHICH IS CLAIMED AND FOR WHICH A PATENT IS SOUGHT ON THE INVENTION

ENTITLED: NANBV DIAGNOSTICS AND VACCINES

the specification of which:

(check one) ☐ is attached hereto;
☒ was filed on 5/18/89 as
Application Serial No. 355,002
and was amended on _____ ;
(if applicable)

I HAVE REVIEWED AND UNDERSTAND THE CONTENTS OF THE ABOVE-IDENTIFIED SPECIFICATION, INCLUDING THE CLAIMS, AS AMENDED BY ANY AMENDMENT REFERRED TO ABOVE;

I ACKNOWLEDGE THE DUTY TO DISCLOSE INFORMATION WHICH IS MATERIAL TO THE EXAMINATION OF THIS APPLICATION IN ACCORDANCE WITH TITLE 37, CODE OF FEDERAL REGULATIONS, Sec. 1.56 (a) which states: "A duty of candor and good faith toward the Patent and Trademark Office rests on the inventor, on each attorney or agent who prepares or prosecutes the application and on every other individual who is substantively involved in the preparation or prosecution of the application and who is associated with the inventor, with the assignee or with anyone to whom there is an obligation to assign the application. All such individuals have a duty to disclose to the Office information they are aware of which is material to the examination of the application. Such information is material where there is a substantial likelihood that a reasonable examiner would consider it important in deciding whether to allow the application to issue as a patent. The duty is commensurate with the degree of involvement in the preparation or prosecution of the application.";

I hereby claim the benefit under Title 35, United States Code, §120 of any United States application(s) listed below, and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States application in the manner provided by the first paragraph of Title 35, United States Code §112, I acknowledge the duty to disclose material information as defined in Title 37, Code of Federal Regulations, §1.56(a) set forth above which occurred between the filing date of the prior application and the national or PCT international filing date of this application:

Application Serial No.	Filing Date	Status (Patented, Pending, Abandoned)
USSN 341,334	20 April 1989	Pending
USSN 325,338	17 March 1989	Pending
PCT/US88 04125	18 November 1988	Pending
USSN 271,450	14 November 1988	Pending
USSN 263,584	26 October 1988	Abandoned
USSN 191,263	6 May 1988	Pending
USSN 161,072	26 February 1988	Pending
USSN 139,886	30 December 1987	Pending
USSN 122,714	18 November 1987	Pending

As to the subject matter of this application which is common to said earlier application, I do not know and do not believe that the same was ever known or used in the United States of America before my or our invention thereof or patented or described in any printed publication in any country before my or our invention thereof or more than one year prior to said earlier application, or in public use or on sale in the United States of America more than one year prior to said earlier application; that said common subject matter has not been patented or made the subject of an inventor's certificate issued before the date of said earlier application in any country foreign to the United States of America on an application filed by me or my legal representatives or assigns more than twelve months prior to said earlier application; and that the earliest application(s) for patent or inventor's certificate on said invention filed by me or my legal representatives or assigns in any country foreign to the United States of America is identified below, as well as all other such applications (if any) filed more than twelve months prior to the filing date of this application:

NA

The priority of the earliest application(s) (if any) filed within a year prior to said pending prior application is hereby claimed under 35 U.S.C. §119;

As to the subject matter of this application which is not common to said earlier application, I do not know and do not believe that the same was ever known or used in the United States of America before my or our invention thereof or patented or described in any printed publication in any country before my or our invention thereof or more than one year prior to the date of this application, or in public use or on sale in the United States of America more than one year prior to the date of this application, and that said subject matter has not been patented or made the subject of an inventor's certificate issued in any country foreign to the United States of America on an application filed by me or my legal representatives or assigns more than twelve months prior to the date of this application, and that the earliest application(s) for patent or inventor's certificate on said subject matter filed by me or my legal representatives or assigns in any country foreign to the United States of America is identified below, as well as all other such application(s) (if any) filed more than twelve months prior to the filing date of this application:

NA

5686836

COMBINED DECLARATION AND POWER OF ATTORNEY FOR CONTINUATION-IN-PART APPLICATION

Attorney Docket No.
2300-0063.28

The priority of the earliest application(s) (if any) filed within a year prior to this application is hereby claimed under 35 U.S.C. §119;

I hereby appoint the following attorneys and agent(s) to prosecute said application and to transact all business in the Patent and Trademark Office connected therewith and to file, prosecute and to transact all business in connection with international applications directed to said invention:

William H. Benz - Reg. No. 25,952
Mary-Elizabeth Buckles - Reg. No. 31,907
Thomas E. Ciotti - Reg. No. 21,013
Ronald Craig Fish - Reg. No. 28,843
Grant D. Green - Reg. No. 31,259
Ronald S. Laurie - Reg. No. 25,431
Gladys H. Monroy - Reg. No. 32,430
Kate H. Murashige - Reg. No. 29,959
Lisabeth Feix Murphy - Reg. No. 31,547
Matthew C. Rainey - Reg. No. 32,291
Dianne E. Reed - Reg. No. 31,292
Roberta Robins - Reg. No. 33,208
Debra Shetka - Reg. No. 33,309

and: Robert P. Blackburn, Reg. No. 30,447

Address all correspondence to:

IRELL & MANELLA
545 Middlefield Road, Suite 200
Menlo Park, California 94025-3471

Address all telephone calls to: Gladys H. Monroy at 415-327-7250.

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

FULL NAME OF SOLE OR FIRST INVENTOR	SIGNATURE	DATE
Michael Houghton	X M Houghton	X 7/13/89
RESIDENCE: Danville, California		CITIZENSHIP: British
POST OFFICE ADDRESS: 53 Rosemead Court, Danville, California 94526		
FULL NAME OF SECOND JOINT INVENTOR, IF ANY	SIGNATURE	DATE
Qui-Lim Choo	X Qui-Lim Choo	X 7/13/89
RESIDENCE: El Cerrito, California		CITIZENSHIP: Singapore
POST OFFICE ADDRESS: 5700 Fern Street, El Cerrito, California 94530		
FULL NAME OF THIRD JOINT INVENTOR, IF ANY	SIGNATURE	DATE
George Kuo	X George Kuo	X 7/13/89
RESIDENCE: San Francisco, California		CITIZENSHIP: U.S.
POST OFFICE ADDRESS: 1370 Sixth Avenue, San Francisco, California 94112		
FULL NAME OF FOURTH JOINT INVENTOR, IF ANY	SIGNATURE	DATE
RESIDENCE		CITIZENSHIP
POST OFFICE ADDRESS		

☐ Please see attached continuation page for additional inventors

5686837-

PTO-103X
(Rev. 11-88)

FILING RECEIPT

UNITED STATES DEPARTMENT OF COMMERCE
Patent and Trademark Office
ASSISTANT SECRETARY AND COMMISSIONER
OF PATENTS AND TRADEMARKS
Washington, D.C. 20231

SERIAL NUMBER	FILING DATE	GRP ART UNIT	FIL FEE REC'D	ATTORNEY DOCKET NO.	DRWGS	TOT CL	IND CL
07/355,002	05/18/89	183	$1,906.00	2300-0063.28	101	40	29

IRELL & MANELLA
545 MIDDLEFIELD RD., STE. 200
MENLO PARK, CA 94025-3471

OCT 6 1989
RECEIVED

Receipt is acknowledged of the patent application identified herein. It will be considered in its order and you will be notified as to the examination thereof. Be sure to give the U.S. SERIAL NUMBER, DATE OF FILING, NAME OF APPLICANT, and TITLE OF INVENTION when inquiring about this application. Fees transmitted by check or draft are subject to collection. Please verify the accuracy of the data presented on this transmittal.

Applicant(s) MICHAEL HOUGHTON, DANVILLE, CA; QUI-LIM CHOO, EL CERRITO, CA; GEORGE KUO, SAN FRANCISCO, CA.

CONTINUING DATA AS CLAIMED BY APPLICANT-
THIS APPLN IS A CIP OF 07/122,714 11/18/87
AND A CIP OF 07/139,886 12/30/87
AND A CIP OF 07/161,072 02/26/88
AND A CIP OF 07/191,263 05/06/88
AND A CIP OF 07/263,584 10/26/88
AND A CIP OF 07/271,450 11/14/88
AND A CIP OF 01/004,125 11/18/88
AND A CIP OF 07/325,338 03/17/89
AND A CIP OF 07/341,334 04/20/89

FOREIGN FILING LICENSE GRANTED 08/16/89
TITLE
NANBV DIAGNOSTICS AND VACCINES

PRELIMINARY CLASS: 424

DOCKETED CHM
2300-0063 28

Related Action Nos. C-88-02
C-88-02974 and C-88-02994

5420526

(see reverse)

CERTIFICATE OF MAILING BY "EXPRESS MAIL"

"Express Mail" Mailing Label No. NB172698658

Date of Deposit 9/15/89

I hereby certify that this paper or fee is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 CFR 1.10 on the date indicated above and is addressed to the Commissioner of Patents and Trademarks, Washington, D.C. 20231.

DOROTHY M. Paul

(Typed or Printed Name of Person Mailing Paper or Fee)

Dorothy M. Paul

(Signature of Person Mailing Paper or Fee)

CH-89
2300-0089
PATENT

NEW HCV ISOLATE

Technical Field

The present invention relates to new isolates of the viral class Hepatitis C, polypeptides, polynucleotides and anti-bodies derived therefrom, as well as the use of such polypeptides, polynucleotides and antibodies in assays (e.g., immunoassays, nucleic acid hybridization assays, etc.) and in the production of viral polypeptides.

Background

Non-A, Non-B hepatitis (NANBH) is a transmissible disease or family of diseases that are believed to be viral-induced, and that are distinguishable from other forms of viral-associated liver diseases, including that caused by the known hepatitis viruses, i.e., hepatitis A virus (HAV), hepatitis B virus (HBV), and delta hepatitis virus (HDV), as well as the hepatitis induced by cytomegalovirus (CMV) or Epstein-Barr virus (EBV). NANBH was first identified in transfused individuals. Transmission from man to chimpanzee and serial passage in chimpanzees provided evidence that NANBH is due to a transmissible infectious agent or agents. Epidemiologic evidence is suggestive that there may be three types of NANBH: the water-borne epidemic type; the blood or

NANBV and/or BB-NANBV from the class of the prototype isolate, HCV1, described by Houghton et al. See, e.g., EPO Pub. No. 318,216 and U.S. Patent App. Serial No. 355,002, filed 19 May 1989 (available in non-U.S. applications claiming priority therefrom), the disclosures of which are incorporated herein by reference. The nucleotide sequence and putative amino acid sequence of HCV1 is shown in Figure 6. The terms HCV, NANBV, and BB-NANBV are used interchangeably herein. As an extension of this terminology, the disease caused by HCV, formerly called NANB hepatitis (NANBH), is called hepatitis C. The terms NANBH and hepatitis C may be used interchangeably herein. The term "HCV", as used herein, denotes a viral species of which pathogenic strains cause NANBH, as well as attenuated strains or defective interfering particles derived therefrom.

HCV is a Flavi-like virus. The morphology and composition of Flavivirus particles are known, and are discussed by Brinton (1986) THE VIRUSES: THE TOGAVIRIDAE AND FLAVIVIRIDAE (Series eds. Fraenkel-Conrat and Wagner, vol eds. Schlesinger and Schlesinger, Plenum Press), p.327-374. Generally, with respect to morphology, Flaviviruses contain a central nucleocapsid surrounded by a lipid bilayer. Virions are spherical and have a diameter of about 40-50 nm. Their cores are about 25-30 nm in diameter. Along the outer surface of the virion envelope are projections that are about 5-10 nm long with terminal knobs about 2 nm in diameter.

The HCV genome is comprised of RNA. It is known that RNA containing viruses have relatively high rates of spontaneous mutation, i.e., reportedly on the order of 10^{-3} to 10^{-4} per incorporated nucleotide. Therefore, there are multiple strains, which may be virulent or avirulent, within the HCV class or species.

JAPANESE JOURNAL
of
MEDICAL SCIENCE & BIOLOGY

Volume 43
1990

PUBLISHED BIMONTHLY BY
THE NATIONAL INSTITUTE OF HEALTH
2-CHOME, KAMIOSAKI, SHINAGAWA-KU
TOKYO 141, JAPAN

Jpn. J. Med. Sci. Biol., 43, 89 - 94, 1990.

Short Communication

DETECTION OF HEPATITIS C VIRUS cDNA SEQUENCE BY THE POLYMERASE CHAIN REACTION IN HEPATOCELLULAR CARCINOMA TISSUES

Tetsuo YONEYAMA, Kenji TAKEUCHI, Yushiro WATANABE[1], Hideharu HARADA[2], Akira OHBAYASHI[2], Yukie TANAKA, Tazuko YUASA, Izumu SAITO and Tatsuo MIYAMURA

Laboratory of Hepatitis Viruses II, Department of Enteroviruses, National Institute of Health, Kamiosaki, Shinagawa-ku, Tokyo 141, [1]Second Department of Internal Medicine, St. Marianna University School of Medicine, 2-16-1 Sugao, Miyamae-ku, Kawasaki-shi, Kanagawa 213, and [2]Liver Unit, Department of Gastroenterology, National Tokyo Sanatorium Hospital, 3-1-1, Takegaoka, Kiyose-shi, Tokyo 204

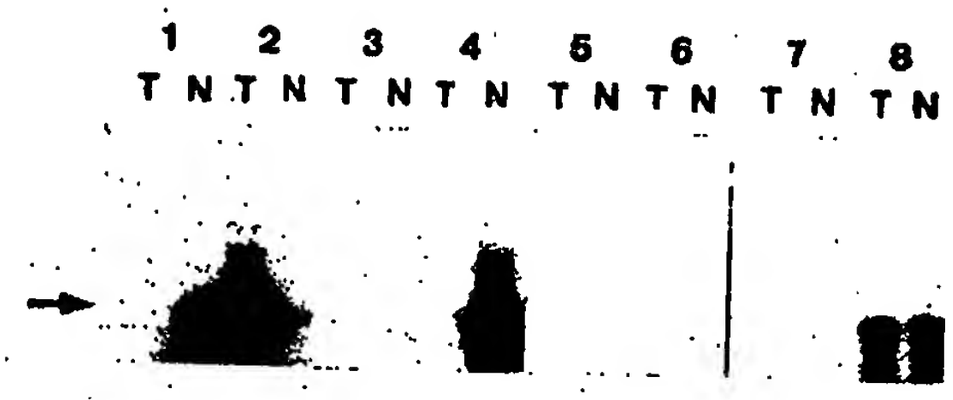
(Received May 16, 1990. Accepted June 28, 1990)

SUMMARY: We found the presence of hepatitis C virus (HCV) infection in liver tissues of hepatocellular carcinoma (HCC) patients who had antibodies to HCV but no serological markers for hepatitis B virus infection by the sensitive reverse transcription/polymerase chain reaction (R/PCR) method. The primers used were derived from the non-structural (NS) 3 and/or the structural (C/E) region. Amplified cDNA sequences of HCV were detected in either cancerous or non-cancerous portion of liver tissues from four out of eight HCC patients with primers of NS3 region. Similar but less efficient results were obtained with primers of C/E region. These results indicate that HCV persists in the liver tissue of HCC. A possible role of persistent infection of HCV for the development of HCC is discussed.

米山徹夫・竹内健司・田中幸江・湯浅田鶴子・斎藤　泉・宮村達男 (国立予防衛生研究所腸内ウイルス部)
渡辺勇四郎 (聖マリアンナ医科大学　川崎市宮前区菅生2-16-1)
原田英治・大林　明 (国立療養所東京病院　清瀬市竹ヶ丘3-1-1)

The cDNA of the HCV genome was recently cloned and identified (1). An assay system for circulating antibodies to HCV was developed by use of an HCV antigen synthesized in recombinant yeast (2). The results obtained from this assay have shown that HCV is the major causative agent of transfusion-associated non-A, non-B hepatitis (2-4), and also a high prevalence of HCV-antibodies among HCC patients who had no serological markers for hepatitis B virus infection (non-B) (5,6). A close association between HCV infection and development of HCC has been suggested (5,6). Furthermore, HCC developed in a chimpanzee after inoculation of human plasma containing non-A, non-B hepatitis agent (7). To study the mechanism of development of HCC, it is necessary to examine the liver tissues of HCC patients for HCV. Recently, cDNA fragments of the HCV genome were cloned from Japanese HCV carriers (8,9). A comparison in nucleotide sequences of Japanese and USA isolates has revealed that there are some heterogeneities of the viral genome between the two isolates. With the nucleotide sequence available, it is now possible to detect the genome of HCV in blood or liver tissues of infected patients (10).

Liver tissues were obtained from autopsy samples of eight non-B HCC patients who had antibodies to HCV. Cancerous and non-cancerous portions were taken from each patient and confirmed histologically. RNA was extracted by the guanidine thiocyanate/cesium fluorotriacetate method (11) from each tissue sample. About 4 μg of RNA was used for cDNA synthesis with 10 units of reverse transcriptase (Bio-Rad, Richmond, CA) and with an antisense primer, J513A (5'-CGTATGAGACACTTCCACAT-3'). The cDNA was amplified by PCR (12) after addition of sense primer, J469S (5'-GTCACTCAGACGGTCGATTT-3'), encompassing the 440 base pairs (bp) of the non-structural protein region 3 (NS) as previously described (8,9). Another set of primers, J135A (5'-ACAGCTTGTGGGATCCGGAG-3') and 71S (5'-GCCGACCTCATGGGGTACAT-3'), which bound the 440 bp of the structural protein region [C/E: putative core and envelope regions (13), K. Takeuchi et al.; manuscript submitted] was also used. Thirty cycles of PCR were carried out as follows: denaturation for 1.5 min at 95 C, annealing of primers for 1.5 min at 55 C, and extension for 2 min at 70 C. The amplified cDNA fragment was electrophoresed on a 2% agarose gel and then transferred to a nylon membrane by a vacuum blotting system (LKB, Bromma, Sweden). A specific signal was identifeid by Southern blot hybridization under the stringent condition with a ^{32}P-labeled probe encompassing target regions of NS3 or C/E. The probe in the



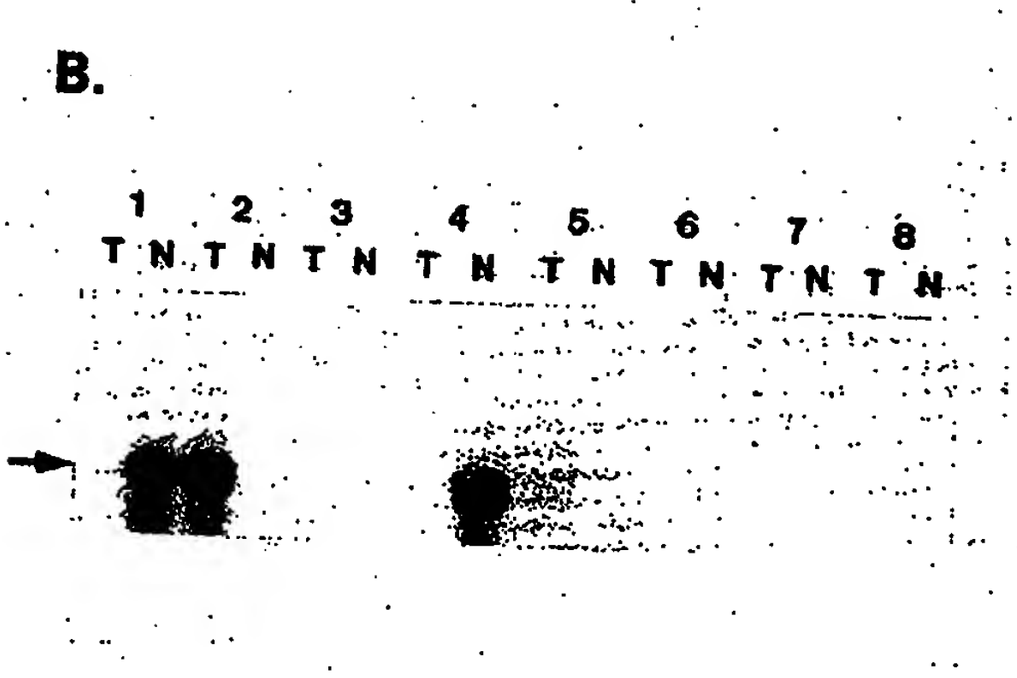


Fig. 1. Detection of cDNA sequences of the HCV genome from liver tissues of HCCs. RNAs were prepared from cancerous (T) and non-cancerous (N) portions of autopsy samples of HCC patients who had antibodies to HCV. Products of R/PCR using either NS3 or C/E primers were electrophoresed on a 2% agarose gel, and analyzed by Southern blot hybridization with a probe encompassing either NS3 (A) or C/E region (B). HCC patients were indicated by numbers 1-8. The arrow indicates the position of the amplified 440 bp cDNA fragment of NS3 or C/E region.

donor of post-transfusion non-A, non-B hepatitis in Japan. Nucleic Acids Res., 17, 10367-10372.
9. Takeuchi, K., Boonmar, S., Kubo, Y., Katayama, T., Ohbayashi, A., Choo, Q.-L., Kuo, G., Houghton, M., Saito, I. and Miyamura, T. (1990): Hepatitis C viral cDNA clones isolated from a healthy carrier donor implicated in post transfusion non-A, non-B hepatitis. Gene, 91, 287-291.
10. Weiner, A. J., Kuo, G., Bradley, D. W., Bonino, F., Saracco, G., Lee, C., Rosenblatt, J., Choo, Q.-L. and Houghton, M. (1990): Detection of hepatitis C viral sequences in non-A, non-B hepatitis. Lancet, 335, 1-3.
11. Okayama, H., Kawaichi, M., Brownstein, M., Lee, F., Yokota, T. and Arai, K. (1987): High-efficiency cloning of full-length cDNA; Construction and screening of cDNA expression libraries for mammalian cells. *In* R. Wu and L. Grossman (eds.), Methods Enzymol., 154, 3-28. Academic Press, New York.
12. Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B and Erlich, H. A. (1988): Primer-derected enzymatic amplification of DNA with a thermostable DNA polymerase. Science, 239, 487-491.
13. Takeuchi, K., Kubo, Y., Boonmar, S., Watanabe, Y., Katayama, T., Choo, Q.-L., Kuo, G., Houghton, M., Saito, I. and Miyamura, T. (1990): Nucleotide sequence of core and envelope genes of the hepatitis C virus genome derived directly from human healthy carriers. Nucl. Acids Res., in press.
14. Rakela, J. and Redeker, A. G. (1979): Chronic liver disease after acute non-A, non-B viral hepatitis. Gastroenterology, 77, 1200-1202.
15. Ohta, Y. (1976): Viral hepatitis and hepatocellular carcinoma. *In* K. Okuda and R. L. Peters (eds.), Hepatocellular Carcinoma, p.73-81. John Wiley & Sons, Inc., New York.
16. Katayama, T., Kikuchi, S., Tanaka, Y., Saito, I., Miyamura, T., Choo, Q.-L., Houghton, M. and Kuo, G. (1990): Blood screening for non-A, non-B hepatitis by hepatitis C virus antibody assay. Transfusion, 30, 374-376.

Proc. Natl. Acad. Sci. USA
Vol. 87, pp. 2057–2061, March 1990
Evolution

Hepatitis C virus shares amino acid sequence similarity with pestiviruses and flaviviruses as well as members of two plant virus supergroups

(non-A, non-B hepatitis/potyvirus/carmovirus/picornavirus/alphavirus)

ROGER H. MILLER AND ROBERT H. PURCELL

Hepatitis Viruses Section, Laboratory of Infectious Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892

Contributed by Robert H. Purcell, December 27, 1989

ABSTRACT **Hepatitis C virus (HCV) is an important human pathogen that is associated with transfusion-related non-A, non-B hepatitis. Recently, HCV cDNA was cloned and the nucleotide sequence of approximately three-quarters of the virus genome was determined. A region of the predicted polyprotein sequence was found to share similarity with a nonstructural protein encoded by dengue virus, a member of the flavivirus family. We report here that HCV shares an even greater degree of protein sequence similarity with members of the pestivirus group (i.e., bovine viral diarrhea virus and hog cholera virus), which are thought to be distantly related to the flaviviruses. In addition, we find that HCV shares significant protein sequence similarity with the polyproteins encoded by members of the picornavirus-like and alphavirus-like plant virus supergroups. These data suggest that HCV may be evolutionarily related to both plant and animal viruses.**

In recent years non-A, non-B (NANB) hepatitis has become the most common form of posttransfusion hepatitis (for reviews, see refs. 1–4). Although first discovered over a decade ago the etiological agent has remained elusive (5, 6). Studies involving the experimental inoculation of chimpanzees provided evidence that the infectious agent was a lipid-containing virus 30–60 nm in diameter bearing strong resemblance to members of the Togaviridae family (7–11). Since titer of the virus in serum rarely reaches 10^6 chimpanzee infectious doses in patients, or experimentally infected animals, additional research has been difficult.

Recently, a λgt11 library was constructed with cDNA synthesized from the RNA of the putative etiological agent of NANB hepatitis (12). Protein synthesized by a specific recombinant reacted exclusively with sera from NANB patients (13). Molecular hybridization analysis demonstrated that the etiological agent, termed hepatitis C virus (HCV), is an RNA virus with a genome size of ≈10 kilobases. The sequence of nearly three-quarters of the virus genome has been reported (14). Analysis indicates that the virus genome is of the plus, or message sense, polarity and appears to lack a poly(A) tail at its 3′ end. The virus genome encodes a single polyprotein, a portion of which shares amino acid sequence similarity with the nonstructural number 3 (NS3) protein of dengue type 2 virus, a member of the flavivirus family. Additional computer-assisted protein analysis, presented here, demonstrates that HCV shares sequence similarity with the polyproteins of animal pestiviruses as well as those of the carmovirus and potyvirus families of plant viruses.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

MATERIALS AND METHODS

Computer Analysis. Computer analysis was through the BIONET National Computer Resource for Molecular Biology. The program FASTA (15) was used to search the European Molecular Biology Organization (EMBO) and GenBank nucleotide data bases and the Swiss (SWS) and National Biomedical Research Foundation (NBRF) protein data bases for sequences with similarity to HCV sequences. FASTA, a derivative of the FASTP program that can be used for both nucleotide and amino acid data base searches, allows multiple regions of similarity between two sequences to be joined to determine a maximum alignment. Briefly, for a protein data base search, an initial similarity score is calculated based on a parameter that determines how many consecutive identities are required in a match and on the total number of identical and similar amino acids as specified by the PAM-250 matrix (16). Next, the FASTA program determines whether several regions with high initial similarity values can be aligned. If so, the program produces an optimal similarity score. There are several limitations imposed when using this program on BIONET. One is that only data base files, and not individual user files, can be analyzed. The second limitation is that only one scoring matrix (i.e., the PAM-250 matrix) can be used for the analysis. Within the FASTA program is a program RDF2 that evaluates the statistical significance of similarity scores by calculating a mean value and the standard deviation from the mean for the similarity scores of sequences in the data base. In this study, a stringent cutoff value for significance of ≥20% amino acid identity in ≥100 residues was also incorporated. Values cited in the text are given as optimized similarity scores with accompanying standard deviation units above the mean calculated for each data base search.

Three programs were used to determine regions of amino acid similarity considering only identical matches in the scoring matrix (17–19). The program HOMOLOGY was used to search for local regions of identity. Residues occurring in the alignments are cited in the text along with the probability that the matches occurred due to chance (e.g., $P = 0.05$ signifies that there is a 5% chance that the same match could occur between random sequences of the same size). The program ALIGN was used to determine the similarity over longer protein domains that encompassed regions with statistically significant matches of identical amino acids. The calculated value H_{max} is directly proportional to the degree of similarity between two sequences over a region of defined size. It should be noted that H_{max} scores produced by the alignment of random sequences range from 20 to 25 for sequences of 190

Abbreviations: HCV, hepatitis C virus; NANB, non-A, non-B; CARMV, carnation mottle virus; NS, nonstructural.

amino acids using the default parameter settings of the program and a segment size of 195 amino acids. Finally, the program GENALIGN was used for multiple sequence alignment.

RESULTS

Houghton *et al.* (14) have reported the nucleotide sequence of approximately three-quarters of the HCV genome. The predicted polyprotein sequence, translated from the NS protein region of the HCV genome, is 2416 amino acids long. Analysis by Houghton and coworkers revealed that, among the virus sequences examined, the polyprotein sequence of HCV was most similar to that of a flavivirus. They reported a similarity between a 530-amino acid domain of the HCV polyprotein sequence and the NS3 protein sequence of dengue virus. We were intrigued by the uniqueness of the HCV sequence and performed searches using several programs to identify global or local regions of significant similarity between HCV and other sequences. This was of special interest since the nucleotide sequences of two pestivirus genomes, bovine viral diarrhea virus (20) and hog cholera virus (21), were determined recently.

First, we used computer-assisted nucleotide sequence analysis to look for similarity between HCV and any sequence recorded in the data base files. Computer searches conducted using the program FASTA with the HCV RNA genome as the query sequence did not result in a statistically significant match with nucleotide sequences in the EMBO or GenBank data bases. These results are in agreement with those of Houghton and coworkers (14). Thus, we conclude that the genome of HCV is not closely related to that of any known RNA virus.

Next, data base searches using the FASTA program and the PAM-250 matrix of Dayhoff (16) were performed to detect protein sequences possessing significant global similarity to the HCV polyprotein. HCV query sequences used were the complete 2416-amino acid polyprotein sequence, as well as the N terminus (i.e., residues 1–1299), and the C terminus (i.e., residues 1200–2416) of the reported HCV genome polyprotein. Searches were conducted using both the SWS and NBRF protein sequence data bases. The FASTA search of the NBRF data base using the entire 2416-residue HCV sequence produced one statistically significant alignment. We found that the amino acid sequence of HCV shared 20.6% amino acid identity with the dengue type 2 virus (22) NS3 protein over a 618-amino acid domain that encompassed the 530-amino acid region of similarity reported by Houghton *et al.* (14). In addition to the 141 matches between identical amino acids, there were 262 amino acids matched by the PAM-250 matrix for a total similarity of 60%. The optimized similarity score of 137 was 11.6 SD units away from the mean value of the analysis. The search of the SWS data base using the 2416-residue HCV polyprotein did not produce a statistically significant alignment. Therefore, using the 2416-amino acid sequence as the query sequence only one alignment score was statistically significant in our analysis.

The FASTA search of both the NBRF and SWS data bases with the N terminus of the HCV polyprotein as the query sequence yielded an alignment that was identical to the one described above. The FASTA search of the two data bases using the C terminus of the HCV polyprotein as the query sequence produced unexpected results. A statistically significant alignment was identified between residues 2058 and 2380 of the HCV polyprotein and the putative replicase of carnation mottle virus (CARMV), a member of the carmovirus group of plant viruses (23). Over a domain of 331 amino acids 67 (20%) of the residues were identical and 126 (38%) were scored as similar by the PAM-250 matrix for a total similarity of 58% (Fig. 1). The optimized similarity score of the alignment was 140, which was 11 SD units above the mean score of the search. Overall, the HCV polyprotein was found to possess significant global similarity to only two sequences in the protein data bases.

```
          2058
HCV       VRCHARKAVTHINSVWKDLLEDNVTPIDTTIMA--KNEVFCVQPEKGGRKPARLIVFPDL
          | |:: :  |  :::  :||: :::: |:::::  |:| | |: :::   ::     |
CARMV     VDCYQGRKRTIYENAAASLLDRAIERKDGDLKTFIKAEKFNVNLKSDPAPRVIQPRSPRY
          356

HCV       GVRVCEKMALYDVVTKLPLAVM--GSSYGFQYSPGQRVEFLVQAWKSKKTPMGFSYDTRC
          :| :  :  |:  :  :|:.:  |::    |:::: :: :  ||:  :||::::::| :
CARMV     NVELGRYLKKYEHHAYKALDKIWGGPTVMKGYTTEEVAQHIWSAWNQFQTPVAIGFDMSR

HCV       FDSTVTESDIRTEEAIYQCCDLDPQARVAIKSLTERLYVGGPLTNSRGENCGYRRCRASG
          ||  |: :::  |:: |  | :: :|::|    : :  |  ::::   :::   || ||
CARMV     FDQHVSVAALEFEHSCYLAC-FEGDAHLANLLKMQLVNHGVGFASNGMLRYTKEGCRMSG
                                                                  *

HCV       VLTTSCGN-TLTCYIKARAACRAAGLQDCTMLVCGDDLVVICESAGVQEDAASLRAFTEA
           ::|: ||  |:| |: :          : ::  ||| |:|||:::::  :::|   |::
CARMV     DMNTALGNCLLACLITKHLM-----KIRSRLINNGDDCVLICERTDIDYVVSNL---TTG
             *  *                          ***

HCV       MTRYSAPPGDPPQPEYDLELITSCSSNVSVAHDGAGKRVYYLTRDPTTPLARAAWETAR-
           :|::     : :| |::| |  |  ::: : ||||   : ::||| :::::::: : ::
CARMV     WSRFGF-NCIAEEPVYEMEKIRFC--QMAPVFDGAG---WLMVRDPLVSMSKDSHSLVHW

                                  2380
HCV       --HTPVNSWLGNIIMFAPTLWARMILMTHFF
            :| :: || :: | :  : : : :: :|:
CARMV     NNETNAKQWLKSVGMCGLRIAGGVPVVQEFY
                                  671
```

FIG. 1. Alignment of the HCV polyprotein sequence (single-letter code) with the putative replicase of CARMV. Residues 2058–2380 of the predicted genome polyprotein of HCV (14) are aligned with residues 356–671 of CARMV (23) that are thought to represent the sequences specifying the virus replicase. Identical amino acid matches are connected with a solid line, while matches scored as similar by the PAM-250 matrix are connected with a colon. Dashes represent spaces between adjacent amino acids that have been inserted to optimize the alignment. Asterisks highlight the six amino acids that have been shown to be invariant among RNA virus replicases (24).

Next, we used several programs to determine whether the HCV polyprotein shared local regions of similarity with other virus sequences scoring only identical amino acid matches. Analysis using the program HOMOLOGY revealed the presence of statistically significant amino acid matches between HCV and two pestivirus polyprotein sequences. For example, the HCV sequences VVLATATPPGSVT (residues 874–886) and QRRGRTGRGKPGIYR (residues 1016–1030) were statistically similar to the bovine viral diarrhea virus (20) sequences VVAMTATPAGSVT (residues 2043–2055) and QRRGRVGRVKPGRYYR (residues 2199–2214) at the $P = 0.007$ and 0.0005 levels, respectively. For reference purposes, we term the former HCV sequence region A and the latter HCV sequence region B. Similar findings were obtained when analyzing the hog cholera virus protein sequence (21). HCV regions A and B were also found to be similar to flavivirus and plant potyvirus polyprotein sequences; however, no such similarity was detected by comparing HCV to alphavirus, rubivirus, or picornavirus protein sequences. For example, the HCV sequence TATPPGS (residues 878–884) in region A was found to be identical to the dengue type 4 virus (25) sequence TATPPGS (residues 1796–1802), which is a statistically significant match at the $P = 0.044$ level. This sequence alignment was also present in the global alignment of Houghton and coworkers (14) and in our alignment using the program FASTA as described above. In addition, the HCV sequence LVVLATATPPG (residues 873–883) of region A was significantly similar to the tickborne encephalitis virus NS3 sequence (26) LVLMTATPPG (residues 1806–1815) at the $P = 0.019$ level of significance. Significant similarity was also found between HCV sequence region B and a plant potyvirus protein sequence. Specifically, the HCV sequence QRRGRTGRGKPG (residues 1016–1027) was similar to the sequence QRFGRVGRNKPG (residues 1463–1474) of the tobacco vein mottling virus (27) at the $P = 0.018$ level of significance. Overall, two regions of the "NS3-like" region of the HCV polyprotein were found to share sequence similarity with pestivirus, flavivirus, and potyvirus proteins.

Table 1. H_{max} similarity values

	HCV	HOG	BVD	TBE	JEV	YFV	DEN	WNF	KUN	TVM
HCV	—	51	52	41	41	40	38	37	33	47
HOG	—	—	169	38	34	33	35	37	33	47
BVD	—	—	—	39	34	34	38	36	37	48
TBE	—	—	—	—	91	87	90	90	93	35
JEV	—	—	—	—	—	88	118	159	163	35
YFV	—	—	—	—	—	—	94	95	90	39
DEN	—	—	—	—	—	—	—	117	121	36
WNF	—	—	—	—	—	—	—	—	177	34
KUN	—	—	—	—	—	—	—	—	—	35
TVM	—	—	—	—	—	—	—	—	—	—

The following virus sequences were used in the analysis: HCV (14); HOG, hog cholera virus (24); BVD, bovine viral diarrhea virus (23); TBE, tickborne encephalitis virus (25); JEV, Japanese encephalitis virus (28); YFV, yellow fever virus (29); DEN, dengue virus (25); WNF, West Nile fever virus (30); KUN, Kunjin virus (31); TVM, tobacco vein mottling virus (26).

To determine the degree of relatedness among HCV and the proteins of the pesti-, flavi-, and potyviruses, we used several programs to analyze a 190-residue domain encompassing HCV regions A and B. In the program ALIGN, the calculated value H_{max} is directly proportional to the degree of similarity between two sequences over a region of defined size. The analysis indicated that the 190-amino acid region of HCV was most similar to that of bovine viral diarrhea virus ($H_{max} = 52$), hog cholera virus ($H_{max} = 51$), and tobacco vein mottling virus ($H_{max} = 47$). Interestingly, HCV shared more similarity with the potyvirus sequence than it did with any of the flavivirus sequences ($H_{max} = 33–41$) examined (Table 1). Multiple sequence alignment of these four sequences using the program GENALIGN demonstrates that there are 25 amino

```
HCV     VVlaTATPPGSVT---vpHPnIEE--valsttGE------ipfyGkaIPleviKGgrhLiFchskkkc
BVD     VVAMTATPAGSVTTTGQKHP-IEEFIAPEVMKGEDLGSqfLDIAGLKIPVdEMKG-NMLVFVPTRNMA
HOG     VVAMTATPAGtVTTTGQKHP-IEEFIAPEVMKGEDLGSeyLDIAGLKIPVeEMK-nNMLVFVPTRNMA
TVM     iikvsATPPGrecdltt-qypvEllIeeqlslrdfvdaqgtDahadvvkkgdnilvyvasynevdqls

HCV     dELAaKLvAlGiNavaYY--rGldvsvipTS-gdvVvVATdAlmtGyT-gDfDsVID-cntCvtqtVd
BVD     VEvAKKLKAKGYNSGYYYSGEDPaNLRVVTSQSPYViVATNAIESGVTLPDLDtVIDTGLKCEKRvRv
HOG     VEaAKKLKAKGYNSGYYYSGEDPsNLRVVaSQSPYVvVATNAIESGVTLPDLDVVVDTGLKCEKRiRl
TVM     kmLnergflvtkvdGrtmklgGveiitkgsSikkhfiVATNiIEnGVTL-DvDVVVDfGLK----vVp

HCV     fSldPtftietitlPqdavsRtQRRGRtGRgKPGi-YR
BVD     SSKiPFIVTGLKRMAVTvGEQAQRRGRVGRVKPGRYYR
HOG     SpKmPFIVTGLKRMAVTiGEQAQRRGRVGRVKPGRYYR
TVM     nldsdnrlvsyckiPislGERiQRfGRVGRnKPGvalR
```

FIG. 2. Multiple sequence alignment of a conserved domain in the genome proteins (single-letter code) of HCV, pestiviruses, and a plant potyvirus. Alignment of the following regions of the genome polyproteins of four viruses are shown: HCV, residues 874–1030 of HCV (14); BVD, residues 2025–2196 of bovine diarrhea virus (20); HOG, residues 1886–2057 of hog cholera virus (21); TVM, residues 1311–1477 of tobacco vein mottling virus (27). Identically matched amino acids between two or more virus proteins are shown as capital letters connected with a straight line. Unmatched amino acids are depicted with lowercase letters. Dashes represent spaces between adjacent amino acids that have been inserted to optimize the alignment. Invariant residues are highlighted with an asterisk.

acids that are invariant among these diverse virus proteins (Fig. 2). Thus, it is likely that this region was conserved in evolution because the protein has an important biological function in virus replication or gene expression.

DISCUSSION

In this study, we used computer-assisted protein analysis to search for sequences with significant similarity to the HCV polyprotein. To identify sequences sharing global similarity, we used a data base searching program that incorporated the PAM-250 matrix to produce alignments consisting of identical and similar amino acid matches. The analysis revealed that the HCV polyprotein possessed statistically significant similarity to only two sequences in the protein data bases. Both sequences were viral in origin. First, the NS3 protein of dengue type 2 virus aligned with a 618-residue domain located near the N terminus of the HCV polyprotein. This represents an extension of nearly 100 amino acids over an alignment reported by Houghton and coworkers (14) that spanned 530 residues within the same region. Second, the putative replicase of CARMV aligned with a region at the C terminus of the HCV polyprotein. This finding was unexpected since CARMV, a member of the carmovirus family, is a plant virus. Overall, the polyprotein of HCV was found to share global similarity with protein sequences encoded by RNA viruses of both animals and plants, which adds support to the hypothesis that there is an evolutionary relationship between these two virus groups.

Analysis in which programs were used to search for regions of local identity of amino acids revealed that regions of the HCV polyprotein aligned with the NS3 protein sequence of flaviviruses and with corresponding regions of the polyproteins of pestiviruses and plant potyviruses. The similarity was the greatest between HCV and pestiviruses. The reason that this similarity was not detected by others previously, or in our data base searches, was that the pestivirus sequences were published only recently and were not in the data bases for analysis. (Therefore, we analyzed the sequences from user files that we created.) Unexpectedly, we did not find significant similarity between the HCV genome protein sequence and the putative replicase of the flaviviruses or pestiviruses.

Comparative analysis of the polyproteins of the members of the flavivirus family reveals that the sequences of the NS proteins are highly conserved (Fig. 3). Multiple sequence alignment of the predicted polyprotein sequences of Japanese encephalitis (28), yellow fever (29), West Nile (30), Kunjin (31), tickborne encephalitis (26), and three dengue virus isolates (25, 32, 33) demonstrates that there are several regions of high amino acid conservation. Within the consensus polyprotein sequence of ≈3400 amino acids there are 21 domains that possess 5 or more consecutive amino acids that are identical in every flavivirus sequence (unpublished data). Eight of these domains are located in the NS3 protein sequence. The 190-amino acid domain of NS3 that shares sequence similarity with HCV contains 3 of these conserved domains. The first is a 7-residue sequence MTATPPG found at the N terminus of the domain. The second is a 5-residue sequence EMGAN near the C terminus. The third is an 8-residue sequence SAAQRRGR located at the extreme C terminus of the domain. Regarding the latter sequence, although the next 3′ residue is variable among flavivirus sequences the following 2 residues are always GR. Our

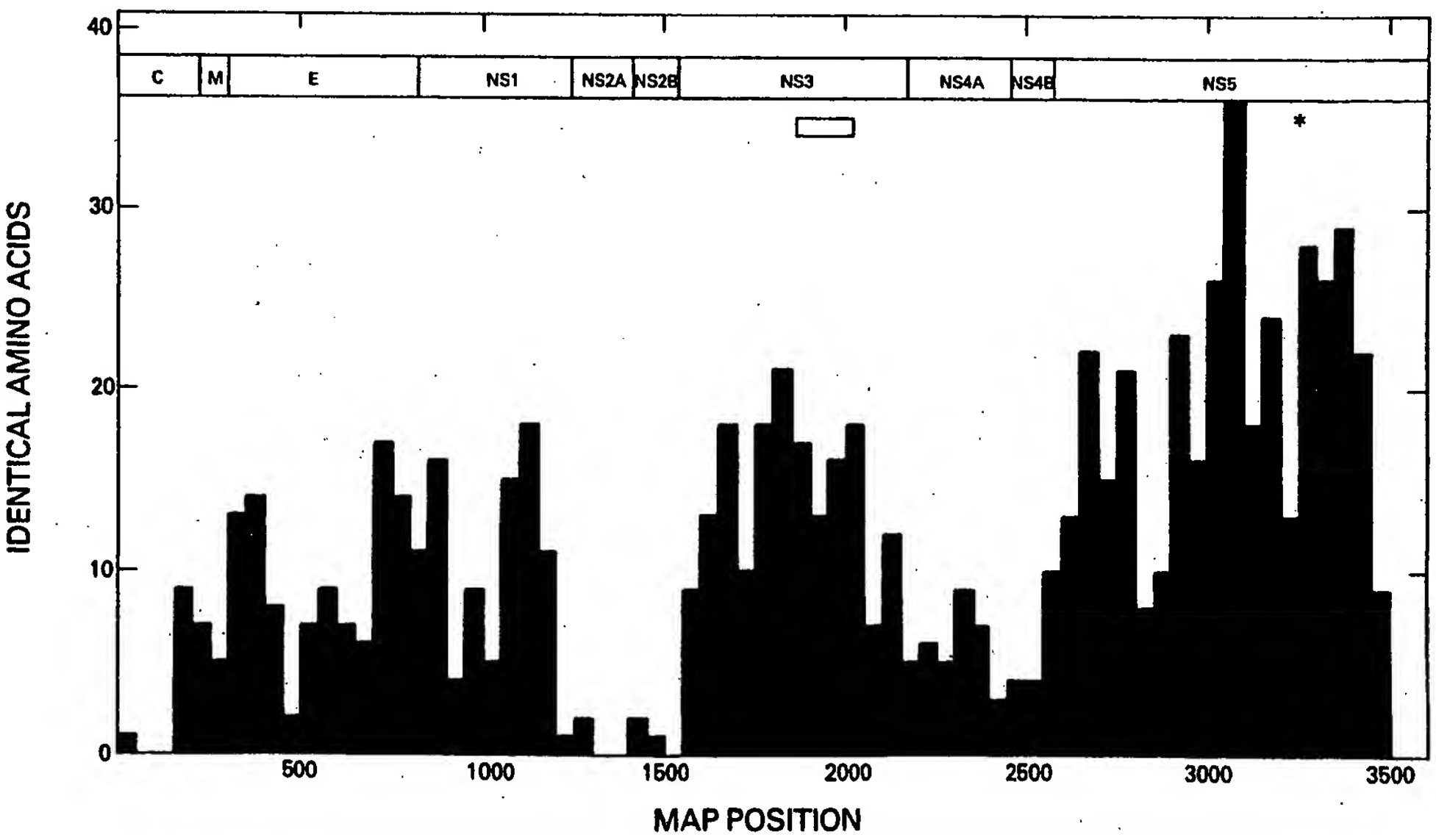


FIG. 3. Histogram of invariant amino acids in the genome polyprotein of the flaviviruses. The program GENALIGN was used to align the amino acids of the following flaviviruses: three isolates of dengue virus (25, 32, 33), Kunjin virus (31), Japanese encephalitis virus (28), tickborne encephalitis virus (26), West Nile virus (30), and yellow fever virus (29). The number of identical amino acids at each position for all 8 sequences, within a block of 50 contiguous residues, is plotted against the position of the residues on the consensus genome polyprotein. The insertion of gaps to optimize the alignment resulted in a total length of the consensus sequence that was longer than any of the individual polyproteins. The gene order of the polyprotein is shown at the top illustrating the position of the structural proteins [i.e., the capsid (C), matrix (M), and envelope (E) proteins] and the NS proteins. The open box under the NS3 protein heading depicts the 190-amino acid domain that shares sequence similarity with regions A and B of the HCV polyprotein. The asterisk represents the position of the invariant GDD moiety of RNA virus replicases.

analysis indicates that only the first and third domains share significant similarity to HCV in the regions of the polyprotein sequence that we have termed A and B.

The NS3 gene region of flaviviruses may encode a protein with several enzymatic activities. First, the N terminus of the NS3 protein is known to share sequence similarity with serine proteases (34). Second, the central domain of NS3 of both flaviviruses and plant potyviruses has been shown to share sequence similarity with helicase-like nucleoside triphosphate binding (NTB) proteins from eukaryotic and prokaryotic cells (35). We find that HCV also shares similarity to NTB proteins in regions A and B of the polyprotein sequence (unpublished data). Thus, it is possible that flavi-, poty-, and pestiviruses, as well as HCV, encode a NTB protein that has been conserved in evolution because of its important catalytic function in virus gene expression or replication.

The NS5 protein has the most highly conserved amino acid sequence of any of the flavivirus proteins and is thought to encode the virus replicase. Within NS5 there are 10 domains that contain ≥5 consecutive identical amino acids including the longest tract of invariant residues (i.e., 14 amino acids) identified in the alignment of the polyproteins. In addition, all flavivirus NS5 proteins possess the 6-amino acid residues that are known to be invariant among RNA polymerase sequences (24). Despite the fact that NS5 is more highly conserved than NS3, we found that there was no statistically significant similarity between the flavivirus NS5 protein and the HCV polyprotein using global or local alignment programs. The only sequence that possessed statistically significant similarity with a region at the C terminus of the HCV polyprotein sequence was the putative replicase of CARMV. Therefore, the HCV replicase may be most closely related to that of a plant virus.

Overall, we find that HCV sequences share significant similarity with proteins from members of two unrelated plant virus families. RNA viruses of plants have been assigned to two supergroups based on the similarity of their genome and protein sequences to either the picorna- or the alphaviruses of animals. The picornavirus supergroup consists of the como-, nepo-, and potyviruses, while the alphavirus, or Sindbis-like, supergroup consists of the alfalfa mosaic, ilar-, bromo-, cucumo-, tobamo-, potex-, tobra-, furo-, nordei-, tombus-, and carmovirus groups (36). There is some speculation that the tombusviruses and carmoviruses may belong to a third supergroup because of their unusually small genome size. The genome of the latter virus group is ≈4000 nucleotides and does not encode an NS3-like protein. Our analysis indicates that amino acid sequences near the N terminus of the HCV polyprotein are similar to those of the potyviruses, while amino acid sequences near the C terminus of the HCV polyprotein are most similar to those of the carmoviruses. Thus, it is possible that HCV represents a recombinant virus possessing an N terminus derived from a picornavirus-like ancestor and a C terminus derived from an alphavirus-like ancestor. However, it is clear that HCV is not closely related to any of these RNA virus families or any other RNA virus family thus far described.

In conclusion, taxonomic classification of HCV must await analysis of the complete nucleotide sequence, which includes the genes encoding the structural proteins as well as the 5′ and 3′ noncoding regions. The data presented here suggest that HCV is distantly related to the pestiviruses and flaviviruses of animals and to members of two plant virus supergroups. It is possible that HCV is a recombinant virus since RNA recombination has been demonstrated for positive-strand (37) and negative-strand RNA viruses (38). Another possibility is that a single virus gave rise to HCV and these similar viruses. Thus, HCV may represent an evolutionary link between the plant virus supergroups and between viruses infecting both plants and animals.

We thank M. Collett and P. Mason for providing pestivirus and flavivirus sequences, respectively, that were not in the data base at the time of the analysis and M. Brinton for helpful discussions. We also thank N. Lofgren, D. Wong, T. Chestnut, and M. Lewis for help in entering nucleotide sequences into computer files and R. Chanock, M. Collett, and M. Brinton for comments on the manuscript. Computer-assisted nucleic acid and protein analysis was through the BIONET National Computer Resource for Molecular Biology, which is supported by National Institutes of Health Grant U41-01685-05.

1. Blum, H. E. & Vyas, G. N. (1982) *Haematologia* **15**, 153–173.
2. Dienstag, J. L. (1983) *Gastroenterology* **85**, 439–462.
3. Fagan, E. A. & Williams, R. (1984) *Semin. Liver Dis.* **4**, 314–335.
4. Mattsson, L. (1989) *Scand. J. Infect. Dis. Suppl.* **59**, 1–55.
5. Feinstone, S. M., Kapikian, A. Z. & Purcell, R. H. (1975) *N. Engl. J. Med.* **292**, 767–770.
6. Alter, H. J., Purcell, R. H., Holland, P. V. & Popper, H. (1978) *Lancet* **i**, 459–463.
7. Feinstone, S. M., Alter, H. J., Dienes, H. P., Shimizu, Y., Popper, H., Blackmore, D., Sly, D., London, W. T. & Purcell, R. H. (1981) *J. Infect. Dis.* **144**, 588–598.
8. Bradley, D. W., McCaustland, K. A., Cook, E. H., Schable, C. A., Ebert, J. W. & Maynard, J. E. (1985) *Gastroenterology* **88**, 773–779.
9. Purcell, R. H., Gerin, J. L., Popper, H., London, W. T., Cicmanec, J., Eichberg, J. W., Newman, J. & Hrinda, M. E. (1985) *Hepatology* **5**, 1091–1099.
10. He, L.-F., Alling, D., Popkin, T., Shapiro, M., Alter, H. J. & Purcell, R. H. (1987) *J. Infect. Dis.* **156**, 636–640.
11. Fagan, E. A., Ellis, D. S., Tovey, G. M., Lloyd, G., Portman, B., Williams, R. & Zuckerman, A. J. (1989) *J. Med. Virol.* **28**, 150–155.
12. Choo, Q.-L., Kuo, G., Weiner, A. J., Overby, L. R., Bradley, D. W. & Houghton, M. (1989) *Science* **244**, 359–362.
13. Kuo, G., Choo, Q.-L., Alter, H. J., Gitnick, G. L., Redeker, A. G., Purcell, R. H., Miyamura, T., Dienstag, J. L., Alter, M. J., Stevens, C. E., Tegtmeier, G. E., Bonino, F., Colombo, M., Lee, W.-S., Kuo, C., Berger, K., Shuster, J. R., Overby, L. R., Bradley, D. W. & Houghton, M. (1989) *Science* **244**, 362–364.
14. Houghton, M., Choo, Q.-L. & Kuo, G. (1988) *Eur. Patent Appl.* 88,310,922.5 and Publ. 318,216.
15. Pearson, W. R. & Lipman, D. J. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 2444–2448.
16. Dayhoff, M., Schwartz, R. M. & Orcutt, B. C. (1978) in *Atlas of Protein Sequence and Structure*, ed. Dayhoff, M. (Natl. Biomed. Res. Found., Silver Spring, MD), Vol. 5, Suppl. 3, pp. 345–352.
17. Needleman, S. B. & Wunsch, C. D. (1970) *J. Mol. Biol.* **48**, 443–453.
18. Korn, L. J., Queen, C. L. & Wegman, M. N. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 4401–4405.
19. Brutlag, D. L., Clayton, J., Friedland, P. & Kedes, L. H. (1982) *Nucleic Acids Res.* **10**, 279–294.
20. Collett, M. S., Larson, R., Gold, C., Strick, D., Anderson, D. K. & Purchio, A. F. (1988) *Virology* **165**, 191–199.
21. Meyers, G., Rumenapf, T. & Thiel, H. (1989) *Virology* **171**, 555–567.
22. Yaegashi, T., Vakharia, V. N., Page, K., Sasaguri, Y., Feighny, R. & Padmanabhan, R. (1986) *Gene* **46**, 257–267.
23. Guilley, H., Carrington, J. C., Balazs, E., Jonard, G., Richards, K. & Morris, T. J. (1985) *Nucleic Acids Res.* **13**, 6663–6677.
24. Domier, L. L., Shaw, J. G. & Rhoads, R. E. (1987) *Virology* **158**, 20–27.
25. Mackow, E., Makino, Y., Zhao, B., Zhang, Y.-M., Markoff, L., Buckler-White, A., Guiler, M., Chanock, R. M. & Lai, C.-J. (1987) *Virology* **159**, 217–228.
26. Mandl, C. W., Heinz, F. X., Stockl, E. & Kunz, C. (1989) *Virology* **173**, 291–301.
27. Domier, L. L., Franklin, K. M., Shahabuddin, M., Hellman, G. M., Overmeyer, J. H., Hiremath, S. T., Siaw, M. F. E., Lomonossoff, G. P., Shaw, J. G. & Rhoads, R. E. (1986) *Nucleic Acids Res.* **14**, 5417–5430.
28. Sumiyoshi, H., Mori, C., Fuke, I., Morita, K., Kuhara, S., Kondou, J., Kikuchi, Y., Nagamatu, H. & Igarashi, A. (1987) *Virology* **161**, 497–510.
29. Rice, C. M., Lenches, E. M., Eddy, S. R., Shin, S. J., Sheets, R. L. & Strauss, J. H. (1985) *Science* **229**, 726–733.
30. Castle, E., Leidner, U., Nowak, T. & Wengler, G. (1986) *Virology* **149**, 10–26.
31. Coia, G., Parker, M. D., Speight, G., Byrne, M. E. & Westaway, E. G. (1988) *J. Gen. Virol.* **69**, 1–21.
32. Deubel, V., Kinney, R. M. & Trent, D. W. (1988) *Virology* **165**, 234–244.
33. Hahn, Y. S., Galler, R., Hunkapiller, T., Dalrymple, J. M., Strauss, J. H. & Strauss, E. G. (1988) *Virology* **162**, 167–180.
34. Gorbalenya, A. E., Donchenko, A. P., Koonin, E. V. & Blinov, V. M. (1989) *Nucleic Acids Res.* **17**, 3889–3897.
35. Lain, S., Riechmann, J. L., Martin, M. T. & Garcia, J. A. (1989) *Gene* **82**, 357–362.
36. Goldbach, R. & Wellink, J. (1988) *Intervirology* **29**, 260–267.
37. Cooper, P. D. (1965) *Virology* **25**, 431–438.
38. Khatchikian, D., Orlich, M. & Rott, R. (1989) *Nature (London)* **340**, 156–157.

JOURNAL OF VIROLOGY, Dec. 1994, p. 8147–8157
0022-538X/94/$04.00+0

Vol. 68, No. 12

Hepatitis C Virus NS3 Serine Proteinase: *trans*-Cleavage Requirements and Processing Kinetics

CHAO LIN, BÉLA M. PRÁGAI, ARASH GRAKOUI, JIAN XU, AND CHARLES M. RICE*
Department of Molecular Microbiology, Washington University School of Medicine, St. Louis, Missouri 63110-1093

Received 1 August 1994/Accepted 16 September 1994

The hepatitis C virus H strain (HCV-H) polyprotein is cleaved to produce at least 10 distinct products, in the order of NH_2-C-E1-E2-p7-NS2-NS3-NS4A-NS4B-NS5A-NS5B-COOH. An HCV-encoded serine proteinase activity in NS3 is required for cleavage at four sites in the nonstructural region (3/4A, 4A/4B, 4B/5A, and 5A/5B). In this report, the HCV-H serine proteinase domain (the N-terminal 181 residues of NS3) was tested for its ability to mediate *trans*-processing at these four sites. By using an NS3-5B substrate with an inactivated serine proteinase domain, *trans*-cleavage was observed at all sites except for the 3/4A site. Deletion of the inactive proteinase domain led to efficient *trans*-processing at the 3/4A site. Smaller NS4A-4B and NS5A-5B substrates were processed efficiently in *trans*; however, cleavage of an NS4B-5A substrate occurred only when the serine proteinase domain was coexpressed with NS4A. Only the N-terminal 35 amino acids of NS4A were required for this activity. Thus, while NS4A appears to be absolutely required for *trans*-cleavage at the 4B/5A site, it is not an essential cofactor for serine proteinase activity. To begin to examine the conservation (or divergence) of serine proteinase-substrate interactions during HCV evolution, we demonstrated that similar *trans*-processing occurred when the proteinase domains and substrates were derived from two different HCV subtypes. These results are encouraging for the development of broadly effective HCV serine proteinase inhibitors as antiviral agents. Finally, the kinetics of processing in the nonstructural region was examined by pulse-chase analysis. NS3-containing precursors were absent, indicating that the 2/3 and 3/4A cleavages occur rapidly. In contrast, processing of the NS4A-5B region appeared to involve multiple pathways, and significant quantities of various polyprotein intermediates were observed. NS5B, the putative RNA polymerase, was found to be significantly less stable than the other mature cleavage products. This instability appeared to be an inherent property of NS5B and did not depend on expression of other viral polypeptides, including the HCV-encoded proteinases.

Hepatitis C viruses (HCVs) have recently been recognized as agents of the parentally transmitted form of non-A, non-B hepatitis (17, 41). Virtual elimination of HCV-contaminated blood has greatly reduced the incidence of posttransfusion hepatitis; however, HCV remains responsible for a significant proportion of community-acquired hepatitis (1). In most cases, HCV is not cleared and establishes a chronic infection that can be associated with chronic hepatitis and more severe liver disease such as cirrhosis and hepatocellular carcinoma (63). For these reasons, there is considerable interest in developing additional HCV-specific antiviral agents that can complement currently available alpha interferon therapy, which effectively controls disease in only a minority of HCV-infected patients.

At least 15 full-length HCV genome sequences, as well as partial sequences for many other isolates, have been reported (see reference 60 and citations therein). These data indicate the existence of multiple genotypes that can diverge by as much as 50% at the amino acid level (10, 64, 65). This group of related viruses is now classified as a separate genus in the family *Flaviviridae* (27), which includes two other genera, *Flavivirus* (12) and *Pestivirus* (20). The positive-strand HCV genome RNA is approximately 9.4 kb in length and contains a highly conserved 5′ noncoding region followed by a long open reading frame encoding a polyprotein of 3,010 to 3,033 amino acids (36, 51). Because a cell culture system supporting efficient HCV replication is lacking, efforts to define potential HCV-encoded polypeptides have utilized expression of HCV cDNA in cell-free translation and in cell cultures. The HCV polyprotein appears to be cleaved at multiple sites to produce at least 10 structural and nonstructural (NS) proteins (47). The order and nomenclature of these cleavage products for the HCV H strain (HCV-H) are NH_2-C-E1-E2-p7-NS2-NS3-NS4A-NS4B-NS5A-NS5B-COOH, where C, E1, and E2 are putative structural proteins and the remaining NS proteins are believed to be replicase components (30–32, 47). Host signal peptidase in the endoplasmic reticulum lumen appears to catalyze cleavages in the structural-NS2 region (C/E1, E1/E2, E2/p7, and p7/NS2 sites) (33, 47), whereas an HCV-encoded serine proteinase located in the N-terminal one-third of the NS3 protein is responsible for four cleavages in the NS region (3/4A, 4A/4B, 4B/5A, and 5A/5B sites) (5, 22, 30, 34, 50, 69). Autocatalytic cleavage at the 2/3 site is mediated by a second HCV-encoded proteinase that encompasses the NS2 region and the NS3 serine proteinase domain (31, 35).

In this study, we tested the ability of the NS3 serine proteinase domain (called $NS3_{181}$) to mediate *trans*-processing at each of the four downstream sites. All four sites could be cleaved in *trans*; however, requirements for *trans*-cleavage varied for different sites. *trans*-cleavage at the 3/4A site was very inefficient, if there was any, when the substrate contained an inactivated serine proteinase. Coexpression of NS4A is required for cleavage at the 4B/5A site, but not at the 5A/5B site. We also tested the ability of the serine proteinases from two HCV subtypes (H and BK strains) to mediate *trans*-processing of heterologous HCV polyprotein substrates. Finally, we used a vaccinia virus recombinant expressing the entire HCV polyprotein to examine the processing kinetics in

* Corresponding author. Mailing address: Department of Molecular Microbiology, Washington University School of Medicine, Box 8230, 660 S. Euclid Ave., St. Louis, MO 63110-1093. Fax: (314) 362-1232. Electronic mail address: rice@borcim.wustl.edu.

the NS region and the stability of HCV precursors and cleavage products.

MATERIALS AND METHODS

Cell cultures. The BHK-21 and CV-1 cell lines were obtained from the American Type Culture Collection, and the BSC-40 cell line (9) was obtained from D. Hruby (Oregon State University). Cell monolayers were grown in Eagle's minimal essential medium (MEM) supplemented with 2 mM L-glutamine, nonessential amino acids, penicillin, streptomycin, and 10% fetal bovine serum (FBS). The A16 subclone of the human hepatoma HepG2 cell line, generously provided by Alan Schwartz (Washington University), was maintained in Dulbecco's modified Eagle medium supplemented with penicillin, streptomycin, and 10% FBS.

Plasmid constructions. Standard recombinant DNA techniques (61) were used for construction of the expression plasmids described below. For all plasmids, regions of HCV-H coding sequence amplified by PCR were verified by DNA sequence analysis.

Synthetic oligonucleotides and PCR were used to engineer initiation or termination codons as well as convenient restriction sites for subcloning (5′ *Nco*I and 3′ *Xho*I sites) for several HCV-H expression constructs. These constructs (with the encoded polyproteins given in parentheses) are as follows (Fig. 1): pTM3/HCV1027-1657 (NS3), pBRTM/HCV1027-1711 (NS3-4A), pTM3/HCV1658-1711 (NS4A), pTM3/HCV1658-1972 (NS4A-4B), pTM3/HCV1658-2420 (NS4A-5A), pTM3/HCV1712-2420 (NS4B-5A), pBRTM/HCV1712-3011 (NS4B-5B), pBRTM/HCV1973-3011 (NS5A-5B), and pTM3/HCV 2421-3011 (Met-NS5B). The sequences encompassing the engineered initiation codons (boldface) are as follows (HCV-H sequence underlined): NS3, 5′-CCATGGCGCCC-3′; NS4A, 5′-CCATGGCCAGCACC-3′; NS4B, 5′-CCATGGCGTCTCAG-3′; NS5A, 5′-CCATGGGATCCGGC-3′; and NS5B, 5′-CCATGGGCTCAATG-3′. For the engineered termination codons (boldface), the surrounding sequences are as follows (HCV-H sequence underlined): NS3, 5′-GTCACGTGACTCGAG-3′; NS4A, 5′-GAGTGCTAGCTCGAG-3′; NS4B, 5′-CCATGCTAGCTCGAG-3′; and NS5A, 5′-TGCTGCTAGCTCGAG-3′.

pTM3/Ubiquitin-HCV2421-3011 (Ubi-NS5B) was constructed by ligation of two PCR-derived fragments into pTM3/HCV2421-3011 (Met-NS5B). The initiating methionine of the ubiquitin monomer corresponds to the ATG in the *Nco*I site of pTM3. The ubiquitin (double underlined)-NS5B (underlined) junction was created by using a *Bam*HI restriction site (boldface) as follows: CGC GGT GGA TCC ATG TCT. The template for PCR amplification of the ubiquitin cassette was pTM3/Ub-nsP4 (Tyr) (44).

Additional HCV-H expression plasmids (with the encoded polyproteins given in parentheses) were constructed by subcloning appropriate fragments from previously described constructs (Fig. 1). pTM3/HCV1027-1207 ($NS3_{181}$) was derived from pTM3/HCV1027-1657 (described above) and pTM3/HCV827-1207 (31); pTM3/HCV1027-1676 ($NS3\text{-}4A_{19}$) was derived from pTM3/HCV1027-1657 and pBRTM/HCV1-1676 (32); pTM3/HCV1027-1692 ($NS3\text{-}4A_{35}$) was derived from pTM3/HCV1027-1657 and pBRTM/HCV1-1692 (32); pBRTM/HCV1027-3011 $S_{1165}A$ (NS3-5B*) was derived from pBRTM/HCV1-3011 $S_{1165}A$ (30) and pBRTM/HCV1027-1711; pTM3/HCV1193-1657 ($NS3_{167\text{-}631}$) was derived from pBRTM/HCV1193-3011 (30) and pTM3/HCV1027-1657; and pTM3/HCV1193-1711 ($NS3_{167}\text{-}4A$) was derived from pBRTM/HCV 1193-3011, pBRTM/HCV1027-1711, and pTM3. pTM3/HCV 1658-1676 ($NS4A_{19}$) was constructed by deleting the *Hin*cII-*Nhe*I fragment of pTM3/HCV1658-1711 (the *Nhe*I site was filled in by using T4 DNA polymerase prior to ligation). pTM3/HCV1658-1692 ($NS4A_{35}$) was generated by deleting the *Nae*I-*Nhe*I fragment of pTM3/HCV1658-1711 (the *Nhe*I site was filled in by using T4 DNA polymerase prior to ligation). pTM3/HCV2269-2508 ($NS5A_{297}\text{-}5B_{88}$) was made by subcloning the 1,274-bp *Bsa*I-*Bgl*II fragment from pTM3/HCV1-2508 (32) into pTM3 digested with *Nco*I and *Bgl*II (the *Bsa*I and *Nco*I sites were filled in by T4 DNA polymerase prior to ligation). pTM3/HCV2285-2508 ($NS5A_{313}\text{-}5B_{88}$) was constructed by subcloning the 1,227-bp *Apa*I-*Bgl*II fragment of pTM3/HCV1-2508 (32) into pTM3, which had been previously

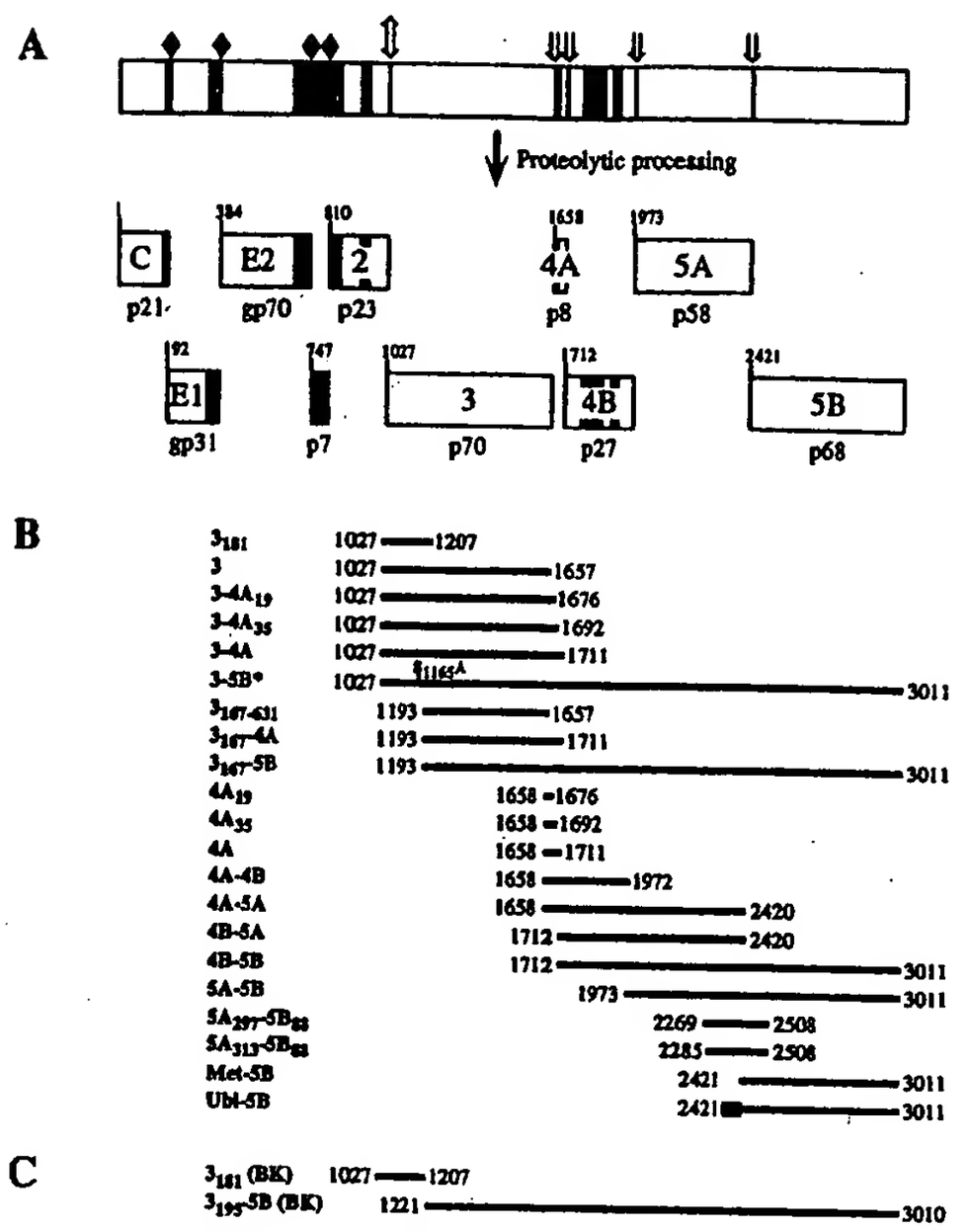


FIG. 1. HCV genome structure and expression constructs. (A) Diagram of the HCV-H strain polyprotein and its cleavage products shown as boxes. The identities of the mature proteins, including C, E1, E2, p7, NS2, NS3, NS4A, NS4B, NS5A, and NS5B, are indicated (32, 47). The number at the top of each cleavage product indicates the position of its N-terminal residue in the polyprotein sequence. The apparent molecular masses for HCV proteins (p) and glycoproteins (gp) are indicated under each product (in kilodaltons). Regions containing predominantly uncharged amino acids are indicated as black bars. Also shown are putative cleavage sites for host signal peptidase (◆) (33, 47), the HCV NS2-3 proteinase (⇳) (31, 34), and the NS3 serine proteinase (⇓) (5, 22, 30, 34, 50, 69). (B) HCV-H polypeptide expression constructs used in this study. HCV polypeptide sequences present in each pBRTM/HCV or pTM3/HCV construct are indicated by black lines, which are drawn to scale and oriented with respect to the diagram of the HCV-H polyprotein. Numbers at the ends of each line refer to the first and last amino acids of the HCV polypeptide expressed by the particular construct. For simplicity, the NS prefix is not used for the nomenclature of each encoded polypeptide, which is indicated on the left. (C) HCV-BK polypeptide expression constructs. (See the legend to panel B for details.)

digested with *Nco*I and *Bgl*II (the *Apa*I and *Nco*I cleavage sites were trimmed and filled in, respectively, by T4 DNA polymerase prior to ligation).

Expression constructs for the HCV BK strain (HCV-BK) were made with cDNA clones generously provided by H. Okayama and A. Takamizawa (67). pTM3/HCV-BK1027-1207 [encoding polypeptide $NS3_{181}$(BK)] was constructed by subcloning a PCR fragment amplified from pUC19/BK-146 (67) into pTM3. The sequences encompassing the engineered initiation and termination codons (boldface) include (HCV-BK sequences underlined) an *Nco*I site at the N terminus (5′-CCATGGCTCCC-3′) and a *Bam*HI site at the C terminus (5′-CGGTCTTAATAGGATCC-3′). pBRTM/HCV-BK1221-3011 was produced by subcloning appropriate fragments from four HCV-BK cDNA clones, including pUC19/BK-102, BK-112-1, BK-112-5, and BK-166. Because the HCV-BK coding sequence in pUC19/BK-102 clones was fused in frame to the AUG codon in the *Nco*I site of the adaptor sequence, pBRTM/HCV-BK1221-3011 encodes a polyprotein [$NS3_{195}$-5B(BK)] encompassing HCV-BK residues 1221 to 3011 after the initiating methionine.

Generation and growth of vaccinia virus-HCV recombinants. vHCV1027-1207 was generated by marker rescue of pTM3/HCV1027-1207 (49). Recombinant viruses were plaque purified three times under *gpt* selection (25) prior to growth of large-scale stocks. A vaccinia virus-HCV recombinant encoding the entire HCV-H open reading frame, vHCV1-3011, has been described previously (47). Stocks of vHCV1027-1207, vHCV1-3011, and vTF7-3, a vaccinia virus recombinant expressing the T7 DNA-dependent RNA polymerase (28), were grown in BSC-40 monolayers and partially purified (37), and titers of infectious progeny were determined by plaque assay on BSC-40 cells (37).

Transient expression with the vaccinia virus-T7 hybrid system. For expression assays utilizing vaccinia virus-HCV recombinants, monolayers of HepG2-A16 or BHK-21 cells in 35-mm-diameter dishes were infected with vTF7-3 alone or in combination with vHCV1-3011, vHCV827-3011 (32), or vHCV1027-1207. The multiplicity of infection for each recombinant was 10 PFU per cell. After adsorption for 60 min at room temperature, the inoculum was removed and replaced with MEM containing 2% FBS. Expression assays of transfected plasmid constructs utilized subconfluent monolayers of BHK-21 cells that had been previously infected with vTF7-3 as described above. Some of them were also coinfected with vHCV1027-1207. After removal of the inoculum, cells were transfected for 2 h at 37°C with a mixture consisting of 1 μg of plasmid DNA and 10 μg of Lipofectin (Bethesda Research Laboratories) in 0.5 ml of MEM. If two constructs were used in a single transfection, the amount of each plasmid varied from 0.5 μg to 1 μg, with a total of 1.5 μg of DNA mixed with 15 μg of Lipofectin.

For pulse-chase experiments, monolayers were washed once with prewarmed methionine-deficient MEM at 3 h postinfection and incubated in the same medium for 20 min at 37°C. Cells were labeled by incubation for 20 min at 37°C with methionine-deficient MEM supplemented with 100 μCi of ^{35}S-protein labeling mixture (NEN) per ml. For chase experiments, the labeling mixture was replaced with MEM containing 2% FBS, 1.5 mg of methionine per ml, and 100 μg of cycloheximide per ml and incubated for the indicated periods at 37°C. For steady-state labeling, cell monolayers were washed once at 3 h postinfection as described above and then were incubated for 4 h at 37°C with MEM containing 1/40th the normal concentration of methionine and cysteine, 2% FBS, and 40 μCi of ^{35}S-protein labeling mixture per ml.

Cell lysis, immunoprecipitation, and protein analyses. After labeling, cell monolayers were washed with phosphate-buffered saline and lysed with a solution of 0.5% sodium dodecyl sulfate (SDS), 50 mM Tris-Cl (pH 7.4), 1 mM EDTA, and 20 μg of phenylmethylsulfonyl fluoride per ml (0.3 ml/10^6 cells). Cellular DNA was sheared by repeated passage through a 27.5-gauge needle. Prior to immunoprecipitation (47), lysates were heated to 70°C for 10 min. Portions of each lysate were incubated either with 5 to 10 μl of the indicated rabbit polyclonal antisera or with 2 μl of serum JHF from an HCV-positive patient (32). Immune complexes were collected by using *Staphylococcus aureus* Cowan I (Calbiochem) as described previously (59), solubilized, and analyzed by SDS-polyacrylamide gel electrophoresis (PAGE) (42) or Tricine-SDS-PAGE (62). After treatment for fluorography with En^{3}Hance (DuPont), gels were dried and exposed at −70°C with prefogged (43) X-ray film (Kodak). ^{14}C-methylated molecular weight marker proteins were purchased from Amersham.

Cell-free translation. The 5′-uncapped RNA transcripts were synthesized from linearized cDNA templates with T7 DNA-dependent RNA polymerase (Epicenter) (58). Cell-free translation mixtures with rabbit reticulocyte lysates (Promega) and [^{35}S]methionine (Amersham), were incubated for 1 h at 30°C essentially according to the manufacturer's instructions. The translation reactions were terminated by the addition of RNase A (Boehringer Mannheim) to 10 μg/ml, cycloheximide to 0.3 mg/ml, and cold methionine to 1 mM. A portion of the translation reaction mixtures was removed at the indicated time, diluted 10-fold with the Laemmli sample buffer, heated for 5 min at 95°C, and analyzed by SDS-PAGE as described above.

RESULTS

***trans*-Cleavage at all four serine proteinase-dependent sites.** The serine proteinase domain of HCVs was initially identified on the basis of sequence homology to members of the trypsin superfamily (7, 29). The predicted domain is approximately 180 residues and corresponds to the N-terminal one-third of NS3. This enzyme is required for processing in the NS3-4-5 region of the HCV polyprotein, and alanine substitutions for predicted active site residues (His-1083 or Ser-1165 for HCV-H) abolish cleavage at the 3/4A, 4A/4B, 4B/5A, and 5A/5B sites (5, 22, 30, 34, 50, 69). To purify and characterize this enzyme, we have used the vaccinia virus-T7 hybrid expression system to examine the ability of the predicted serine proteinase domain, expressed as an individual polypeptide ($NS3_{181}$), to mediate *trans*-cleavage at each of these four sites.

The first substrate examined was an NS3-5B polyprotein containing the Ala substitution at Ser-1165 (NS3-5B*) (Fig. 1). This mutation completely inactivates the serine proteinase, and no processed products were observed (Fig. 2). When coexpressed with $NS3_{181}$, cleavage occurred at the 4A/4B, 4B/5A, and 5A/5B sites, as evidenced by the appearance of NS4B (Fig. 2B), NS5A (Fig. 2C), and NS5B (Fig. 2D). In contrast, we observed a more-slowly-migrating NS3-specific product, presumably NS3-4A, in addition to a very faint band corresponding to NS3 (Fig. 2A). This suggests that very inefficient, if any, *trans*-cleavage occurred at the 3/4A site of this substrate.

The lack of *trans*-cleavage at the 3/4A site has been observed in other studies and has led to the proposal that this site can only be cleaved in *cis* (5, 69). However, all substrates examined thus far contained an inactivated NS3 serine proteinase domain, which might interfere with the accessibility of the 3/4A

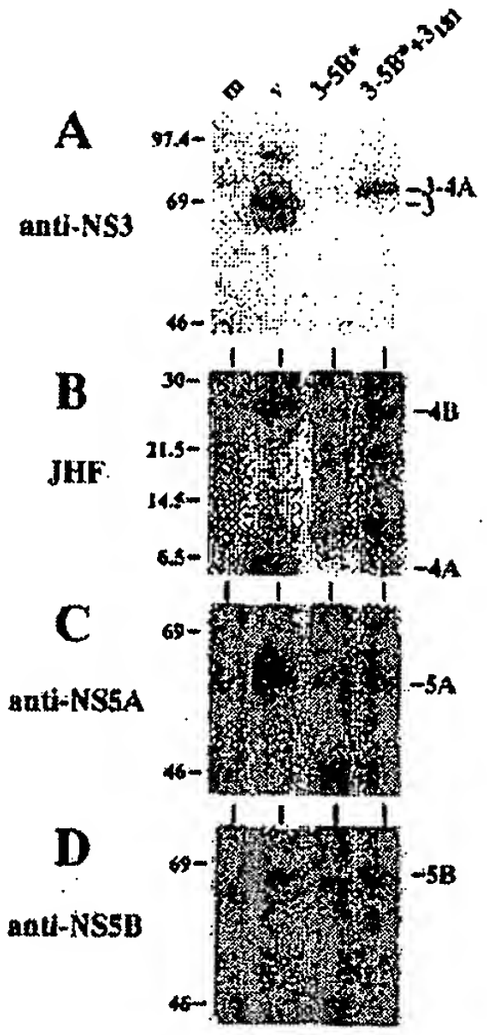


FIG. 2. *trans*-Processing of the HCV-H NS3-5B* polyprotein. BHK-21 cell monolayers were infected with vTF7-3 alone (m) or in combination with vHCV827-3011 (v) or vHCV1027-1207 (3_{181}). Some monolayers were also transfected with pBRTM/HCV1027-3011 $S_{1165}A$ (3-5B*). Cells were metabolically labeled with ^{35}S-protein labeling mixture as described in Materials and Methods. Cell lysates were immunoprecipitated with the following HCV-specific antisera: NS3-specific WU117 (A), NS5A-specific WU123 (C), NS5B-specific WU115 (D), or human patient serum JHF (B). It should be noted that the NS3 serine proteinase domain is not recognized by either human patient serum JHF or rabbit antiserum WU117, which was raised against the NS3 helicase domain. Immunoprecipitated proteins were solubilized and separated by electrophoresis on 8% (A, C, and D) or 14% (B) polyacrylamide–SDS gels. HCV-specific proteins are indicated on the right, and the sizes of ^{14}C-labeled protein molecular mass markers (in kilodaltons) are indicated on the left.

site for *trans*-cleavage. To test this possibility, we expressed a polyprotein, $NS3_{167}$-5B, which begins with residue 167 of NS3 and therefore lacks the majority of the serine proteinase domain. Marker proteins were also expressed beginning with NS3 residue 167 and extending to the C terminus of NS3 ($NS3_{167-631}$) or NS4A ($NS3_{167}$-4A) (Fig. 1). Processed products were not observed when $NS3_{167}$-5B was expressed alone (Fig. 3). During coexpression with $NS3_{181}$, two NS3-specific cleavage products were observed: a major product comigrating with $NS3_{167-631}$ and traces of a larger species comigrating with $NS3_{167}$-4A (Fig. 3). These results clearly demonstrate that $NS3_{181}$ can mediate efficient *trans*-cleavage at the 3/4A site of a substrate which lacks the inactivated proteinase domain.

In contrast to the flaviviruses, where NS2B is absolutely required for NS3 serine proteinase activity (11, 24, 57), HCV sequences upstream of NS3 are not required for serine proteinase-dependent cleavages (5, 22, 30). However, the potential role of downstream viral polypeptide sequences in proteolysis has not been examined. To address this possibility, we tested *trans*-cleavage of NS4A-4B, NS4B-5A, and NS5A-5B substrates, each of which contained only a single proteinase-dependent cleavage site (Fig. 1). When expressed alone, only

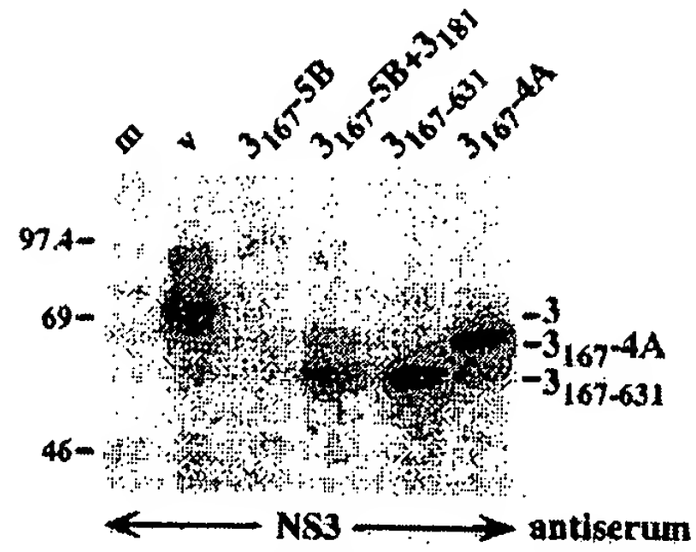


FIG. 3. Requirements for *trans*-cleavage at the 3/4A site. BHK-21 cell monolayers were infected with vTF7-3 alone (m) or in combination with vHCV827-3011 (v) or vHCV1027-1207 (3_{181}). Some monolayers were also transfected with pBRTM/HCV1193-3011 (3_{167}-5B), pTM3/HCV1193-1657 ($3_{167-631}$), or pTM3/HCV1193-1711 (3_{167}-4A). Cells were labeled with ^{35}S-protein labeling mixture as described in Materials and Methods. HCV NS3-specific products were immunoprecipitated with rabbit antiserum WU117, solubilized, and analyzed by SDS-PAGE (8% polyacrylamide). HCV-specific proteins are indicated on the right, and the sizes of ^{14}C-labeled protein molecular mass markers (in kilodaltons) are indicated on the left.

the appropriate unprocessed polyproteins were present (Fig. 4). When coexpressed with $NS3_{181}$, NS4A-4B was processed to yield NS4A and NS4B (Fig. 4A), and NS5A-5B yielded NS5A and NS5B (Fig. 4C). To develop shorter substrates convenient for in vitro proteinase assays, we examined *trans*-processing of $NS5A_{297}$-$5B_{88}$ and $NS5A_{313}$-$5B_{88}$, which contain the C-terminal 152 and 136 residues of NS5A, respectively, followed by the N-terminal 88 amino acids of NS5B (Fig. 1). $NS5_{297}$-$5B_{88}$ was processed efficiently by $NS3_{181}$ as evidenced by the conversion of most of $NS5A_{297}$-$5B_{88}$ to $NS5A_{297-448}$. Nearly complete *trans*-cleavage at the 5A/5B site was also observed for $NS5A_{313}$-$5B_{88}$ (Fig. 4D). These results indicate that only limited flanking sequences are necessary for efficient *trans*-cleavage at the 5A/5B site by the $NS3_{181}$ serine proteinase. Since these substrates do not overlap, these data exclude an absolute requirement for one of the downstream viral polypeptides for serine proteinase activity. In contrast to the results with the NS4A-4B, NS5A-5B, and NS3-5B* substrates, however, no *trans*-cleavage of NS4B-5A was observed (Fig. 4B).

NS4A is required for cleavage at the 4B/5A site. Since *trans*-cleavage at the 4B/5A site occurred for the NS3-5B* substrate but not for NS4B-5A, we examined *trans*-cleavage of NS4A-5A and NS4B-5B polyprotein substrates (Fig. 1). When coexpressed with $NS3_{181}$, the 4B/5A cleavage occurred in the NS4A-5A substrate (Fig. 5A, lane 4) but not in NS4B-5B (data not shown). These results suggested that, in addition to the $NS3_{181}$ proteinase domain, NS4A was required for cleavage at the 4B/5A site. The requirement for NS4A was strengthened by the observation that processing of NS4B-5A was restored by coexpression of NS4A in *trans*. Cleavage at the 4B/5A site of this substrate occurred when NS4A was expressed either as part of the proteinase (NS3-4A) (Fig. 5A, lane 9) or as an individual polypeptide (NS4A) together with $NS3_{181}$ (Fig. 5A, lane 7). These results clearly demonstrate that NS3-mediated cleavage at the 4B/5A site requires NS4A, a small protein of 54 amino acids with a hydrophobic N-terminal half and a C-terminal half rich in charged residues (see Discussion).

Since NS3-4A was fully active for *trans*-cleavage at the 4B/5A site, we made two constructs with C-terminal deletions

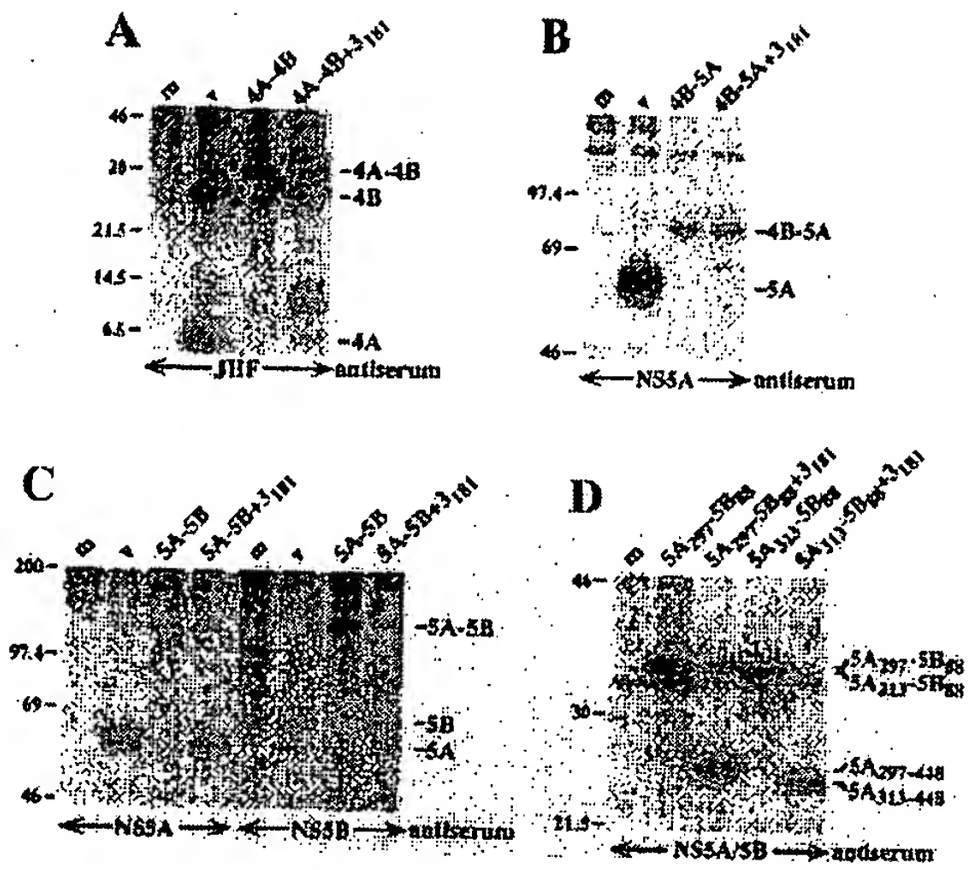


FIG. 4. *trans*-Processing of HCV-H polyproteins containing only one serine proteinase-dependent site. BHK-21 cell monolayers were infected with vTF7-3 alone (m) or in combination with vHCV827-3011 (v) or vHCV1027-1207 (3_{181}). As indicated, some monolayers were also transfected with pTM3/HCV1658-1972 (4A-4B), pTM3/HCV1712-2420 (4B-5A), pBRTM/HCV1973-3011 (5A-5B), pTM3/HCV2269-2508 ($5A_{297}$-$5B_{88}$), or pTM3/HCV2285-2508 ($5A_{313}$-$5B_{88}$). These BHK-21 cells were labeled with ^{35}S-protein labeling mixture as described in Materials and Methods. Cell lysates were immunoprecipitated with human patient serum JHF (A) or the following HCV-specific rabbit antisera: NS5A-specific WU123 (B and C), NS5B-specific WU115 (C), and WU113 specific for both NS5A and NS5B (D). Apparently, rabbit antiserum WU113, which was raised against a fusion protein containing the C-terminal 109 residues of NS5A and the N-terminal 203 residues of NS5B, recognizes only the NS5A region but not the NS5B sequences in $5A_{297}$-$5B_{88}$ and $5A_{313}$-$5B_{88}$. Immunoprecipitated proteins were solubilized and separated by electrophoresis on 14% (A and D) or 8% (B and C) polyacrylamide–SDS gels. HCV-specific proteins are indicated on the right, and the sizes of ^{14}C-labeled protein molecular mass markers (in kilodaltons) are indicated on the left. In panel A, NS4A is difficult to visualize because it contains only a single methionine residue (compared with six in NS4B) and migrates as a diffuse band on this gel system.

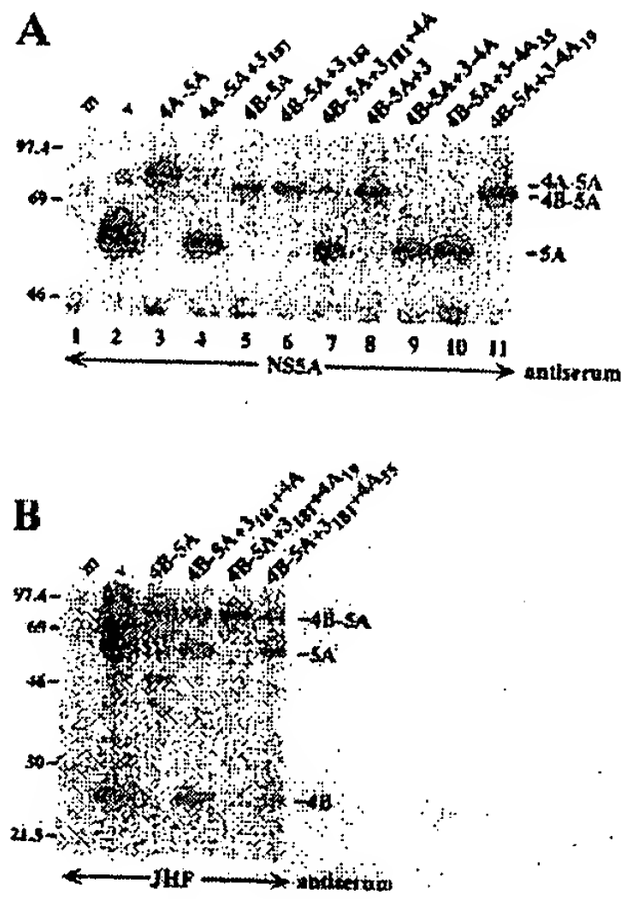


FIG. 5. Requirements for *trans*-cleavage at the 4B/5A site. BHK-21 cell monolayers were infected with vTF7-3 alone (m) or in combination with vHCV827-3011 (v) or vHCV1027-1207 (3_{181}). As indicated, some monolayers were also transfected with the following plasmids: pTM3/HCV1658-2420 (4A-5A), pTM3/HCV1712-2420 (4B-5A), pTM 3/HCV1658-1711 (4A), pTM3/HCV1027-1657 (3), pTM3/HCV1027-1676 (3-$4A_{19}$), pTM3/HCV1027-1692 (3-$4A_{35}$), pBRTM/HCV1027-1711 (3-4A), pTM3/HCV1658-1676 ($4A_{19}$), and pTM3/HCV1658-1692 ($4A_{35}$). Cells were labeled with ^{35}S-protein labeling mixture as described in Materials and Methods. HCV-specific products were immunoprecipitated with NS5A-specific antiserum WU123 (A) or human patient serum JHF (B), solubilized, and separated by 8% (A) or 10% (B) polyacrylamide–SDS gels. HCV-specific proteins are indicated on the right, and the sizes of ^{14}C-labeled protein molecular mass markers (in kilodaltons) are indicated on the left.

in the NS4A region to map NS4A sequences required for this activity. NS3-$4A_{35}$ and NS3-$4A_{19}$ contain the full-length NS3 followed by the N-terminal 35 and 19 residues of NS4A, respectively (Fig. 1). As evidenced by production of NS5A, NS3-$4A_{35}$ (Fig. 5A, lane 10), but not NS3-$4A_{19}$ (lane 11), was able to process NS4B-5A. In an earlier study, similar constructs were generated to map the location of NS4A (32). A polyprotein beginning with the C protein and extending through the N-terminal 35 residues of NS4A was efficiently processed at the 3/4A site. However, a C-terminal truncation to residue 19 of NS4A appeared to block cleavage at the 3/4A site (32). Thus, the inability of NS3-$4A_{19}$ to function for *trans*-cleavage of NS4B-5A might result from lack of cleavage at the 3/4A site and release of the NS4A N terminus rather than deletion of NS4A residues 20 to 35. To address this possibility, we examined the activity of polypeptides encompassing the N-terminal 19 and 35 residues of NS4A (called $NS4A_{19}$ and $NS4A_{35}$, respectively). $NS4A_{35}$, but not $NS4A_{19}$, was able to induce *trans*-cleavage of NS4B-5A by $NS3_{181}$ (Fig. 5B). These results indicate that the C-terminal 19 amino acids (residues 36 to 54) of NS4A, which contain 8 to 9 highly conserved, charged residues (see Discussion), are not required for *trans*-cleavage at the 4B/5A site.

Cleavage at the 3/4A and 4A/4B sites, which flank NS4A, may also require NS4A sequences for efficient cleavage (see Discussion). However, since the 5A/5B site can be efficiently cleaved in the absence of NS4A (Fig. 4C and D), this protein is not absolutely required for NS3 serine proteinase activity. For development of an in vitro proteinase assay that does not require NS4A, substrates containing the 5A/5B site should be good candidates.

***trans*-Cleavage between HCV-H and HCV-BK strains.** Viral proteinases, which are important for polyprotein processing and viral replication, present attractive targets for development of antiviral therapeutic agents. Since sequence analysis of HCV isolates has uncovered considerable genetic diversity, the success of a proteinase inhibitor strategy will depend at least in part on the conservation of proteinase-substrate interactions among different HCV types. In one classification scheme, six major genotypes or types (from 1 to 6) are distinguished, with some types further divided into related subtypes (64, 65). HCV-H (26) and HCV-BK (67) are members of the 1a and 1b subtypes, respectively, which represent the major subtypes in the United States and Japan. These two strains share 90 and 87% amino acid sequence identities in the serine proteinase domain and in the NS3-5B region, respectively. To examine the conservation or divergence of proteinase-substrate interactions among different HCV strains, we compared the ability

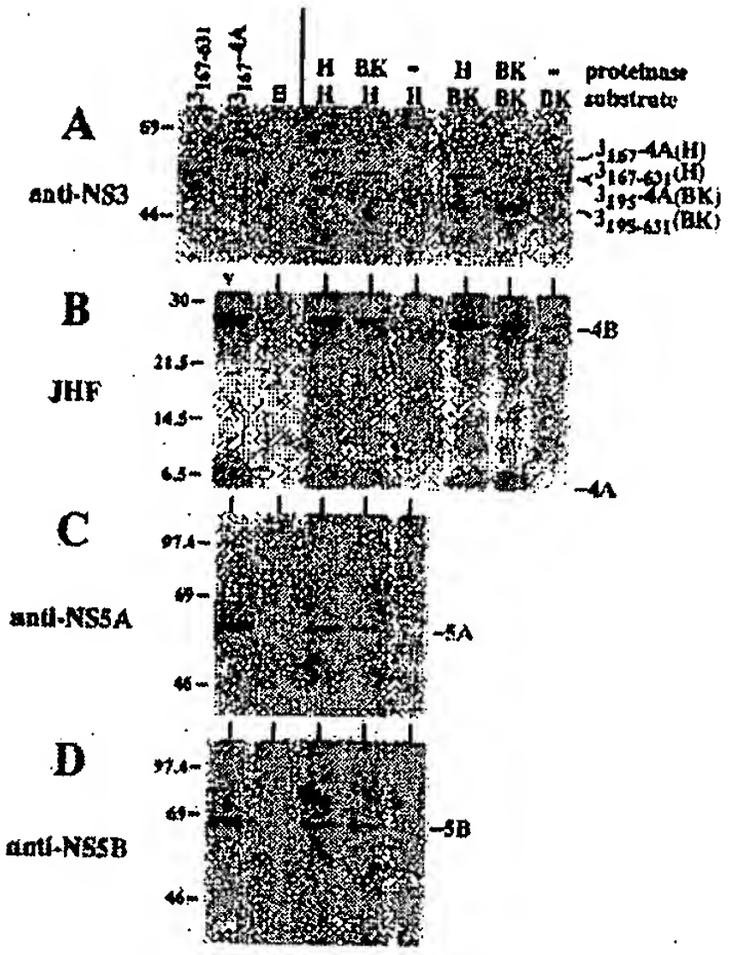


FIG. 6. *trans*-Cleavage between HCV-H and HCV-BK polypeptides. BHK-21 cell monolayers were infected with vTF7-3 alone (m) or in combination with vHCV827-3011 (v). For *trans*-cleavage experiments, the polyprotein substrates were pBRTM/HCV1193-3011 for the H strain (H) or pBRTM/HCV-BK1221-3010 for the BK strain (BK). The serine proteinase domains of both strains were expressed as the source of the proteolytic activities: vHCV1027-1207 for the H strain and pTM3/HCV-BK1027-1207 for the BK strain. The absence (−) of certain expression constructs is indicated. Cells were labeled with ^{35}S-protein labeling mixture as described in Materials and Methods. HCV-specific products were immunoprecipitated with the following antisera: NS3-specific WU117 (A), NS5A-specific WU123 (C), NS5B-specific WU115 (D), or human patient serum JHF (B). The immunoprecipitated proteins were solubilized and separated on 8% (A, C, and D) or 14% (B) polyacrylamide–SDS gels. HCV-specific proteins are indicated on the right, and the sizes of ^{14}C-labeled protein molecular mass markers (in kilodaltons) are indicated on the left.

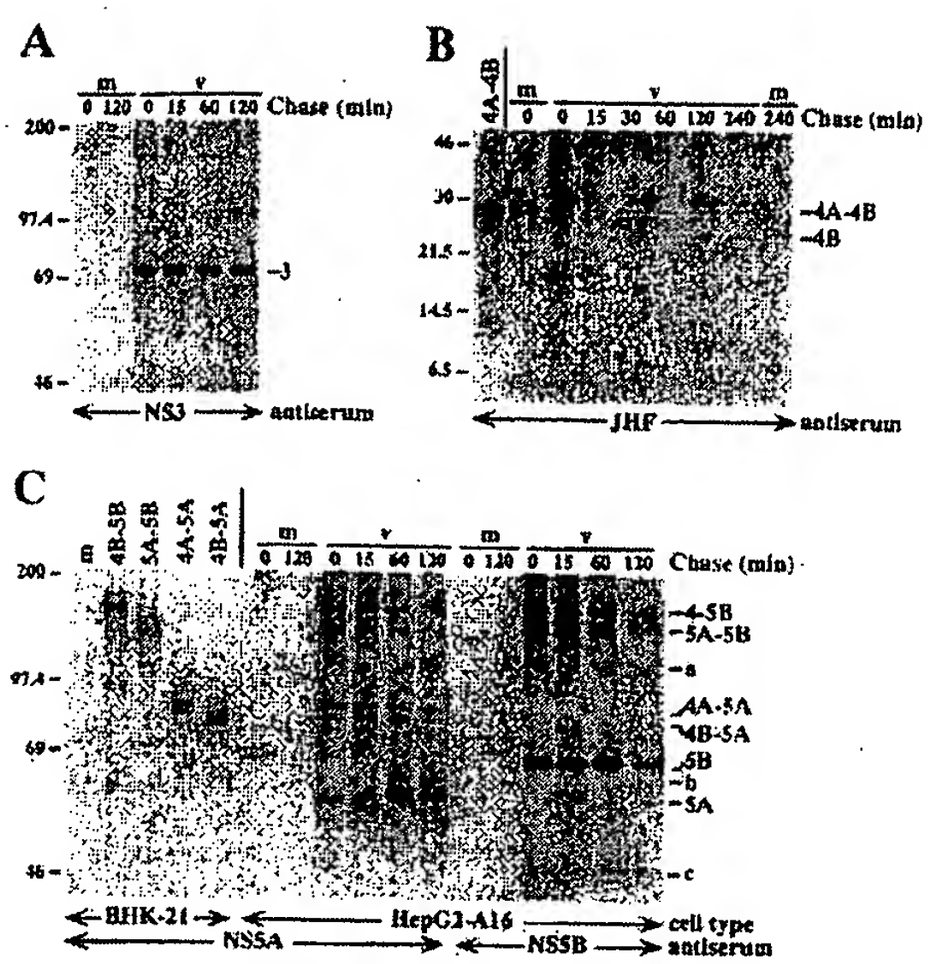


FIG. 7. Pulse-chase analysis of processing in the NS3-4-5 regions. HepG2-A16 cells were infected with vTF7-3 alone (m) or coinfected with vTF7-3 and vHCV1-3011 (v), pulse labeled with ^{35}S-protein labeling mixture for 20 min, and chased for the indicated time as described in Materials and Methods. Cell lysates were prepared and immunoprecipitated with the following antisera: NS3-specific WU117 (A), NS5A-specific WU123, or NS5B-specific WU115 (C) or human patient serum JHF (B). Immunoprecipitated proteins were solubilized and separated by SDS-PAGE (8% polyacrylamide) (A and C) or Tricine-SDS-PAGE (14% polyacrylamide) (B). The migration pattern of HCV NS5A-specific polyprotein markers is shown on the left in panel C. BHK-21 cells previously infected with vTF7-3 were mock transfected (m) or transfected with the indicated plasmids and labeled with ^{35}S-protein labeling mixture as described in Materials and Methods. HCV-specific proteins are identified on the right, and the sizes of ^{14}C-labeled protein molecular mass markers (in kilodaltons) are indicated on the left.

of the NS3 serine proteinases of the HCV-H or the HCV-BK strain to mediate *trans*-cleavage of homologous or heterologous polyprotein substrates.

For the H strain, we used the $NS3_{181}$ proteinase and the $NS3_{167}$-5B substrate described above (Fig. 1). For the HCV-BK proteinase, we made a similar construct expressing the N-terminal 181 amino acids of HCV-BK NS3 [$NS3_{181}$(BK)]. The HCV-BK substrate was a polyprotein beginning with residue 195 of NS3 and extending through NS5B [$NS3_{195}$-5B(BK)]. When $NS3_{167}$-5B was coexpressed with $NS3_{181}$, processing at all four sites occurred, as evidenced by the appearance of $NS3_{167-631}$, NS4A, NS4B, NS5A, and NS5B (Fig. 6). For the BK strain, $NS3_{181}$(BK) was able to mediate *trans*-cleavage at the 3/4A, 4A/4B, and 4B/5A sites of $NS3_{195}$-5B(BK), as indicated by the production of $NS3_{195-631}$(BK), NS4A, and NS4B (Fig. 6A and B). Thus far, we have been unable to identify the HCV-BK NS5A and NS5B cleavage products by using HCV-H NS5A- or NS5B-specific rabbit antisera or HCV-positive patient antisera collected in the United States. As shown in Fig. 6, the serine proteinase domain of either strain was fully active at mediating *trans*-cleavage of the heterologous substrate from the other strain. $NS3_{181}$(BK) cleaved $NS3_{167}$-5B of H strain to $NS3_{167-631}$, NS4A, NS4B, NS5A, and NS5B (Fig. 6). Likewise, $NS3_{195}$-5B(BK) was processed by $NS3_{181}$ of the H strain to produce $NS3_{195-631}$(BK), NS4A, and NS4B (Fig. 6A and B). Thus, at least as assessed by this *trans*-processing assay, these two different HCV subtypes do not appear to have diverged significantly in terms of serine proteinase-substrate recognition.

Kinetics of processing in the HCV NS region. Besides defining the minimal domains required for serine proteinase activity, it is also of interest to understand the processing reactions that occur in the full-length HCV polyprotein. In other viral systems, polyprotein cleavages that occur in *cis* versus those occurring in *trans* can be important for regulating RNA replicase function. Such regulation is possible when polyproteins, processing intermediates, and mature cleavage products have distinct roles in replication (44). To begin to examine processing pathways and kinetics in the NS3-4-5 region, pulse-chase experiments were carried out in HepG2-A16 cells by using a vaccinia virus-HCV recombinant, vHCV1-3011, which expresses the entire HCV-H polyprotein (47). As shown in Fig. 7A, NS3 was readily visible after a 20-min pulse and was not associated with any higher-molecular-mass polyprotein precursors, indicating that cleavage at both the 2/3 and 3/4A sites occurs very rapidly, possibly in *cis*.

In contrast, processing in the NS4-5 region was generally

slower and a number of processing intermediates were readily identified. As shown in Fig. 7B, NS4B was readily visible after a 20-min pulse. A 29-kDa protein comigrating with the product expressed from pTM3/HCV1658-1972 (NS4A-4B) was identified as the NS4A-4B polyprotein (Fig. 7B). A decrease in the level of NS4A-4B was accompanied by an increase in the amount of NS4B, which suggests that NS4A-4B can be a precursor for NS4B and NS4A. NS4A was not observed in this experiment, probably because of its low methionine content (only one) and inefficient expression by vHCV1-3011 (32). Four predominant NS5A-containing polyproteins of 160, 135, 87, and 82 kDa were observed after the 20-min pulse (Fig. 7C). NS4-specific antiserum recognized all of these species except for the 135-kDa polyprotein (data not shown), whereas the NS5B-specific antiserum recognized only the 160- and 135-kDa polyproteins (Fig. 7C). On the basis of their apparent molecular mass, immunoreactivity, and comigration with marker polyproteins (Fig. 7C), these four polyproteins were tentatively identified as NS4-5B (160 kDa), NS5A-5B (135 kDa), NS4A-5A (87 kDa), and NS4B-5A (82 kDa). It is unclear whether the 160-kDa polyprotein NS4-5B begins with NS4A or NS4B or is a mixture of both of these species. The presence of these four polyproteins suggests that there are several alternative pathways for processing the NS4-5 region (see Discussion for more details). Over a 60-min chase, the level of NS5A (58 kDa) increased significantly and was accompanied by a decrease in the levels of NS5A-5B and NS4-5B (Fig. 7C), suggesting that these two polyproteins may be the precursors to NS5A. Because the levels of NS4B-5A and NS4A-5A increased initially, and then decreased during the chase period (Fig. 7C), they probably represent processing intermediates between NS4-5B and NS5A. An NS5A-specific protein of 62 kDa (indicated as b in Fig. 7C), barely detectable after 15 min of chase, became more apparent after 60 min. In a previous study, several minor NS5A-specific species with slower mobility were observed in addition to the dominant 58-kDa NS5A protein (32). Two additional faint bands of 107 and 47 kDa (labeled a and c, respectively, in Fig. 7C) were observed with the NS5B-specific antiserum. Product a was also recognized by NS5A-specific antiserum. These two proteins remain to be defined, but they may reflect additional proteolytic processing within the NS5B region. Although NS4-5B and NS5A-5B were likely precursors to NS5B, the level of NS5B did not change significantly over a 60-min chase period (Fig. 7C), and this protein appeared to be unstable relative to most of the other HCV-encoded proteins (see below). On the other hand, NS3 (Fig. 7A) and NS4B (Fig. 7B) were stable up to 2 h, while a slight decrease in the level of NS5A was observed (Fig. 7C).

Instability of the NS5B protein. While NS3 was very stable during prolonged chase periods, NS5A and, in particular, NS5B, the putative HCV RNA polymerase, appeared to be rather unstable. NS5A disappeared with a half-life of approximately 170 min, and NS5B disappeared with a half-life of about 70 min (data not shown). This observation is potentially interesting because some positive-strand viruses tightly regulate the level of their RNA-dependent RNA polymerase. Additionally, the p75 protein of bovine viral diarrhea virus, the HCV NS5B homolog, is unstable in bovine viral diarrhea virus-infected cells (21). To determine whether other HCV-encoded proteins might be responsible for the instability of NS5B, we expressed two different forms of HCV NS5B. One form (Met-NS5B) included the entire NS5B region preceded by two non-HCV residues, Met-Gly. A second construct encoded a ubiquitin fusion protein consisting of the 76-residue ubiquitin monomer fused in frame to the N terminus of NS5B (Ubi-NS5B). Cleavage of this ubiquitin fusion protein by

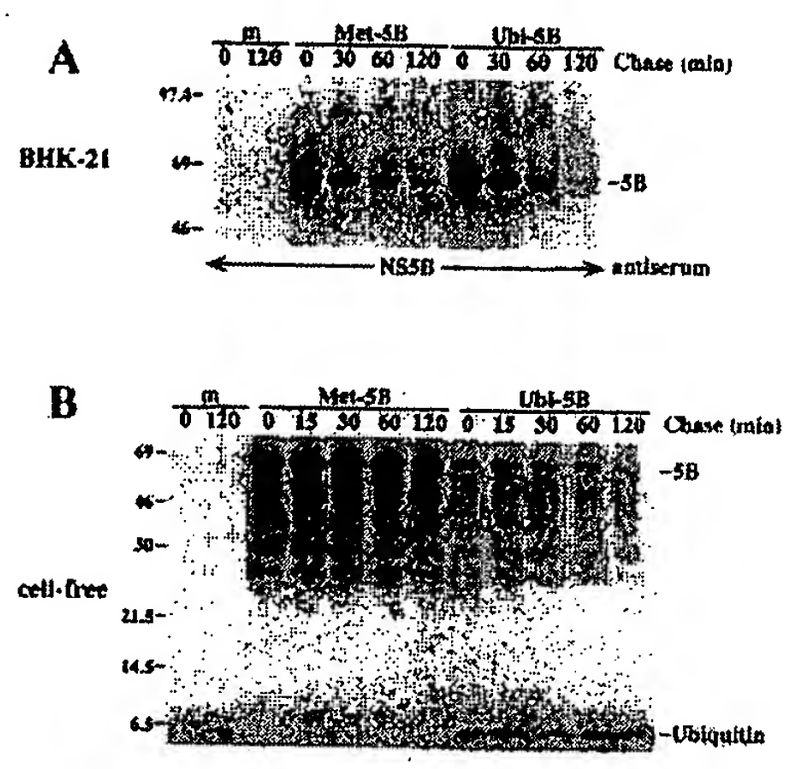


FIG. 8. Stability of NS5B expressed in BHK-21 cells or by cell-free translation. (A) BHK-21 cells previously infected with vTF7-3 were mock transfected (m) or transfected with one of the following plasmids: pTM3/HCV2421-3011 (Met-5B) or pTM3/Ubiquitin-HCV2421-3011 (Ubi-5B). The cell monolayers were pulse-labeled with ^{35}S-protein labeling mixture for 20 min and chased for the indicated times as described in Materials and Methods. Cell lysates were prepared, and HCV NS5B-specific products were immunoprecipitated by the rabbit antiserum WU115. Immunoprecipitated proteins were solubilized and separated by SDS-PAGE (8% polyacrylamide). HCV-specific proteins are identified on the right, and the sizes of ^{14}C-labeled protein molecular mass markers (in kilodaltons) from Amersham are indicated on the left. (B) Translations with RNA transcripts from pTM3/HCV2421-3011 or pTM3/Ubiquitin-HCV2421-3011 or without any transcript (m) were incubated for 60 min at 30°C in reticulocyte lysate in the presence of [^{35}S]methionine. Translation reactions were terminated by the addition of RNase A, cycloheximide, and excess cold methionine and then were chased for the indicated times. The translation products were solubilized and analyzed by SDS-PAGE (14% polyacrylamide). The identities of proteins are shown on the right, and the sizes of ^{14}C-labeled protein molecular mass markers (in kilodaltons) are indicated on the left.

cellular ubiquitin carboxy-terminal hydrolase should produce NS5B with its authentic N-terminal Ser residue (4). As shown in Fig. 8A, the Ubi-NS5B fusion protein was completely processed to NS5B after a 20-min pulse of transfected BHK-21 cells. Both forms of the NS5B proteins were unstable, as evidenced by the rapid decline in the level of NS5B (Fig. 8A). The approximate half-lives were 90 min for Met-NS5B and 70 min for NS5B produced by cleavage of Ubi-NS5B. Coexpression of $NS3_{181}$ had no significant effect on the stability of NS5B (data not shown). These results indicate that the instability of NS5B is not due to the presence of other HCV-encoded proteinases or proteins. Rather, NS5B is inherently unstable and is probably degraded through a cellular pathway.

In an attempt to devise an in vitro assay to study NS5B degradation, we examined the stability of Met-NS5B or Ubi-NS5B produced by cell-free translation of RNA transcripts in rabbit reticulocyte lysates (Fig. 8B). Translation reactions were terminated by addition of RNase A, cycloheximide, and excess cold methionine and were chased for the indicated periods. Control experiments showed no further incorporation of [^{35}S]methionine after the addition of these three reagents (data not shown). Ubi-NS5B fusion proteins were completely processed to NS5B and ubiquitin (predicted molecular mass of 8.6 kDa) after a 60-min incubation. Although slight decreases

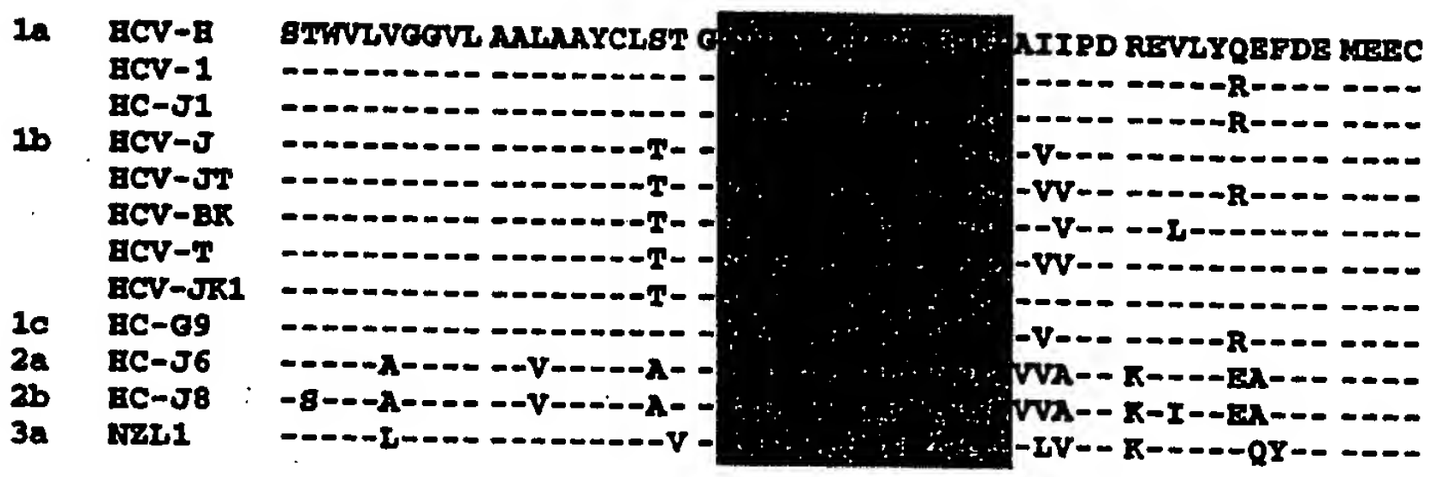


FIG. 9. Alignment of NS4A sequences. The predicted NS4A amino acid sequences are aligned for selected HCV isolates from six subtypes (indicated on the left): HCV-H (38), HCV-1 (18), HC-J1 (accession no. D10749), HCV-J (39), HCV-JT (68), HCV-BK (67), HCV-T (16), HCV-JK1 (accession no. S18030), HC-G9 (53), HC-J6 (55), HC-J8 (54), and HCV-NZL1 (60). The single-letter code for amino acids is used. Hyphens indicate residues identical to those of the HCV-H strain sequence. The 14-residue segment (residues 22 to 35) implicated in *trans*-cleavage at the 4B/5A site is shaded. Accession numbers for unpublished sequences are given above in parentheses.

in the levels of the proteins were apparent over the 2-h chase, both Met-NS5B and NS5B produced by cleavage of Ubi-NS5B were quite stable in the reticulocyte lysates, making it difficult to assess the role of the ubiquitin-mediated degradation pathway (which is present in reticulocyte lysates [19]) in the turnover of NS5B.

DISCUSSION

It has been previously shown that an active NS3 serine proteinase is required for processing at four cleavage sites in the HCV NS3-4-5 region. The results presented here clearly demonstrate that the proteinase domain, expressed as a 181-residue N-terminal fragment of NS3, is able to mediate *trans*-cleavage at all four sites. Bartenschlager et al. (6) recently reported similar results showing that a fragment of the polyprotein, including the 212 N-terminal residues of NS3 and a 20-residue extension into the NS2 region, could also mediate *trans*-cleavage at the 4A/4B, 4B/5A, and 5A/5B sites. Our results, as well as those of two recent studies (6, 23), indicate that NS4A is absolutely required for the 4B/5A cleavage. Failla et al. (23) also showed that NS4A of HCV-BK, supplied in *trans*, was required for cleavage at the 3/4A and 4B/5A sites and improves the efficiency of processing at the 4A/4B and 5A/5B sites. On the basis of these results, it was suggested that NS4A functions as a general effector or cofactor for NS3 serine proteinase-mediated cleavage in the NS3-4-5 region. Virus-encoded cofactors required for serine proteinase activity have also been found for other members of the family *Flaviviridae*. The most dramatic example is the NS2B protein of flaviviruses, which is absolutely required for NS3 serine proteinase-mediated cleavage at all structural and nonstructural dibasic sites (2, 11, 15, 24, 45, 48, 56, 57, 70, 72). As discussed by Failla et al. (23), the pestivirus p10 protein may be the functional homolog of HCV NS4A, because sequences in this region of the pestivirus polyprotein appear to be required for the serine proteinase-dependent cleavage between p58 and p75 (the two C-terminal products of the pestivirus polyprotein possibly equivalent to HCV NS5A and NS5B, respectively) (71). For HCV, NS4A is required for only three cleavages mediated by the serine proteinase (3/4A, 4A/4B, and 4B/5A). While Failla et al. showed that NS4A can increase *trans*-cleavage efficiency at the 5A/5B site (23), we found that certain substrates containing this site could be processed efficiently in the absence of NS4A. Hence, the HCV serine proteinase-dependent cleavages can be separated into at least two types: (i) cleavages at the 3/4A, 4A/4B, and 4B/5A sites, which are located adjacent to hydrophobic sequences and require NS4A as a cofactor; and (ii) cleavage at the 5A/5B site, which can occur in the absence of NS4A.

Although the mechanism(s) by which NS4A functions in proteolytic processing at type 1 sites remains to be determined, several possibilities can be envisioned. (i) NS4A may act as a molecular chaperone to facilitate folding of the serine proteinase domain into an active enzyme. If the active form of the proteinase is the same for cleavage at both type 1 and type 2 sites, then this model implies that type 1 substrates are suboptimal and require higher concentrations of active proteinase for efficient *trans*-cleavage. (ii) NS4A may bind to type 1 substrates, the proteinase domain, or both to facilitate proteinase-substrate interactions and cleavage. (iii) NS4A may facilitate proteolysis of membrane-associated type 1 substrates by interacting with the proteinase domain and localizing it to the membrane compartment. Given that NS4A is required for cleavage at three different sites, it is tempting to propose that it functions via direct interaction with the proteinase domain. Thus far, unlike the flavivirus proteinase, which consists of a stable complex of NS2B and NS3 (3, 14), there is no direct evidence for association between the HCV NS3 and NS4A proteins. Suggestive evidence has been obtained, however, from in vitro studies in which NS3 was found to become membrane associated when the cell-free translation product included the NS4A region (35).

Although NS4A is only 54 residues in length, we showed that a fragment of only 35 N-terminal residues, coexpressed with the serine proteinase domain, was sufficient for *trans*-cleavage at the 4B/5A site. Failla et al. (23) reported that a polypeptide consisting of the C-terminal 33 residues of NS4A and NS4B facilitated *trans*-cleavage at the 4B/5A site. Although flanking sequences may contribute to NS4A activity, these data suggest that a 14-residue segment (residues 22 to 35) of NS4A may be critical for cleavage at the 4B/5A site. As shown in Fig. 9, the HCV NS4A protein sequence is highly conserved among HCV strains and consists of a hydrophobic N-terminal portion, a central region implicated in *trans*-cleavage at the 4B/5A site (highlighted in Fig. 9), and a highly charged acidic C-terminal segment. Although somewhat less conserved than other regions of NS4A (especially in comparison with HCV-J6 and HCV-J8), this central region contains two positively-charged residues, several hydrophobic amino acids, and an absolutely conserved Gly at position 27. The importance of these residues for NS4A *trans*-cleavage activity is currently being tested by site-directed mutagenesis.

The lack of *trans*-cleavage at the 3/4A site in previous studies led to the suggestion that cleavage at this site occurred in *cis* (5, 69). This cleavage has recently been shown to be insensitive to dilution, providing direct evidence for a *cis* mechanism (6). These observations are consistent with the results of pulse-chase analyses in which we (this report) and others (6) were unable to detect NS3-related precursors. Thus, in the current model, both the 2/3 and 3/4A cleavages are catalyzed by two distinct viral proteinases in *cis*. Of interest is the observation that substrates with an inactivated serine proteinase domain were resistant to *trans*-cleavage at the 3/4A site. Efficient *trans*-cleavage was observed, however, when the inactivated proteinase domain was deleted. Although other possibilities exist, these results, together with the observation that NS4A sequences are required for cleavage at the 3/4A site (23), suggest that during translation of the polyprotein, the serine proteinase domain interacts with nascent NS4A to assume a conformation capable of *cis*-cleavage at the 3/4A site. In the case of the substrate with the inactivated proteinase, this intermediate still forms but is frozen because it is inactive for *cis*-cleavage. Thus, the 3/4A site of this substrate is not accessible to *trans*-acting proteinase, because it is probably bound in the substrate binding pocket of the inactive autoproteinase.

In contrast to the rapid cleavages observed at the 2/3 and 3/4A sites, processing at the 4A/4B, 4B/5A, and 5A/5B sites was slower and appeared to involve multiple pathways (this study and reference 6). An obligate processing order was not observed, which is consistent with results from a study in which mutations blocking cleavage at each of these three sites had no significant effect on processing at other sites (40). Similar results have been obtained for flaviviruses (45, 46, 52, 56). It is important to emphasize that the processing pathways and kinetics observed in mammalian transient expression assays may not accurately reflect the situation in HCV-infected cells. In particular, *trans*-processing reactions, which are important for temporal regulation of RNA synthesis for other viruses (for example, see reference 44), would be expected to be sensitive to the concentration of *trans*-acting factors, which may be much lower in HCV-infected cells. Hence, these issues should be reexamined when systems become available for studying HCV replication in cell cultures.

Using both the vaccinia virus-T7 and the Sindbis virus replicon expression systems, we found that the NS5B protein was unstable compared with the other polyprotein cleavage products (8) (Fig. 7 and 8 and data not shown). Turnover of NS5B was similar whether the protein was expressed as part of the full-length polyprotein or independently as a ubiquitin fusion protein. In contrast to the results in cell culture assays, NS5B was found to be relatively stable in reticulocyte lysates. Since NS5B is the putative HCV RNA-dependent RNA polymerase (51), down-regulation of this protein could play an important regulatory role in virus replication, as has been found for the RNA polymerase of alphaviruses (see reference 66 for a review). For the pestivirus bovine viral diarrhea virus, the putative RNA-dependent RNA polymerase (p75) is unstable, with a half-life of less than 60 min in bovine viral diarrhea virus-infected cells (21). However, the NS5 protein of flaviviruses, which is not cleaved into two proteins, is relatively stable (13). In contrast to our results, Bartenschlager et al. found NS5B (of a strain similar to HCV-J) to be quite stable when expressed in HeLa cells with a vaccinia virus recombinant (6). The reason for the discrepancy is unclear, but it could reflect a difference in the sequence of the expressed NS5B protein or in the cells used for the expression studies. As mentioned above, these issues need to be reexamined in a system that supports HCV RNA replication.

Finally, there is considerable interest in the HCV proteinases as targets for development of new antivirus therapies. The general usefulness of such compounds will depend in part on their ability to inhibit the proteinases of diverse HCV types. Although it will be important to test more divergent proteinase-substrate combinations, the ability of the HCV-H serine proteinase (subtype 1a) to *trans*-process an HCV-BK substrate (subtype 1b), and vice versa, suggests that the essential elements of recognition may be conserved. This is encouraging for the development of broadly effective serine proteinase inhibitors.

ACKNOWLEDGMENTS

We thank Bernie Moss for providing pTM3 and vTF7-3; Hiroto Okayama and Akihisa Takamizawa for HCV-BK cDNA clones; Henry Hsu and Harry Greenberg for human patient serum JHF; and Sean Amberg, Julie Lemm, Sasha Kolykhalov, and Karen Reed for critical reading of the manuscript.

This work was supported by a grant from the Public Health Service (CA57973). C.L. is a predoctoral candidate and was supported in part by the Division of Biology and Biomedical Sciences at Washington University. B.M.P. was supported in part by the Ministry of Welfare, T076-ETT (90-93), Hungary.

REFERENCES

1. **Alter, M. J., S. C. Hadler, F. N. Judson, A. Mares, W. J. Alexander, P. Y. Hu, J. K. Miller, L. A. Moyer, H. A. Fields, D. W. Bradley, and H. S. Margolis.** 1990. Risk factors for acute non-A, non-B hepatitis in the United States and association with hepatitis C virus infection. JAMA 264:2231–2235.
2. **Amberg, S. M., A. Nestorowicz, D. W. McCourt, and C. M. Rice.** 1994. NS2B-3 proteinase-mediated processing in the yellow fever virus structural region: in vitro and in vivo studies. J. Virol. 68:3794–3802.
3. **Arias, C. F., F. Preugschat, and J. H. Strauss.** 1993. Dengue 2 virus NS2B and NS3 form a stable complex that can cleave NS3 within the helicase domain. Virology 193:888–899.
4. **Bachmair, A., D. Finley, and A. Varshavsky.** 1986. *In vivo* half-life of a protein is a function of its amino-terminal residue. Science 234:179–186.
5. **Bartenschlager, R., L. Ahlborn-Laake, J. Mous, and H. Jacobsen.** 1993. Nonstructural protein 3 of the hepatitis C virus encodes a serine-type proteinase required for cleavage at the NS3/4 and NS4/5 junctions. J. Virol. 67:3835–3844.
6. **Bartenschlager, R., L. Ahlborn-Laake, J. Mous, and H. Jacobsen.** 1994. Kinetic and structural analyses of hepatitis C virus polyprotein processing. J. Virol. 68:5045–5055.
7. **Bazan, J. F., and R. J. Fletterick.** 1990. Structural and catalytic models of trypsin-like viral proteases. Semin. Virol. 1:311–322.
8. **Bredenbeek, P. J., I. Frolov, C. M. Rice, and S. Schlesinger.** 1993. Sindbis virus expression vectors: packaging of RNA replicons by using defective helper RNAs. J. Virol. 67:6439–6446.
9. **Brockman, W. W., and D. Nathans.** 1974. The isolation of simian virus 40 variants with specifically altered genomes. Proc. Natl. Acad. Sci. USA 71:942–946.
10. **Bukh, J., R. H. Purcell, and R. H. Miller.** 1993. At least 12 genotypes of hepatitis C virus predicted by sequence analysis of the putative E1 gene of isolates collected worldwide. Proc. Natl. Acad. Sci. USA 90:8234–8238.
11. **Chambers, T. J., A. Grakoui, and C. M. Rice.** 1991. Processing of the yellow fever virus nonstructural polyprotein: a catalytically active NS3 proteinase domain and NS2B are required for cleavages at dibasic sites. J. Virol. 65:6042–6050.
12. **Chambers, T. J., C. S. Hahn, R. Galler, and C. M. Rice.** 1990. Flavivirus genome organization, expression, and replication. Annu. Rev. Microbiol. 44:649–688.
13. **Chambers, T. J., D. W. McCourt, and C. M. Rice.** 1990. Production of yellow fever virus proteins in infected cells: identification of discrete polyprotein species and analysis of cleavage kinetics using

region-specific polyclonal antisera. Virology 177:159–174.
14. **Chambers, T. J., A. Nestorowicz, S. M. Amberg, and C. M. Rice.** 1993. Mutagenesis of the yellow fever virus NS2B protein: effects on proteolytic processing, NS2B-NS3 complex formation, and viral replication. J. Virol. 67:6797–6807.
15. **Chambers, T. J., R. C. Weir, A. Grakoui, D. W. McCourt, J. F. Bazan, R. J. Fletterick, and C. M. Rice.** 1990. Evidence that the N-terminal domain of nonstructural protein NS3 from yellow fever virus is a serine protease responsible for site-specific cleavages in the viral polyprotein. Proc. Natl. Acad. Sci. USA 87:8898–8902.
16. **Chen, P.-J., M.-H. Lin, K.-F. Tai, P.-C. Liu, C.-J. Lin, and D.-S. Chen.** 1992. The Taiwanese hepatitis C virus genome: sequence determination and mapping the 5′ termini of viral genomic and antigenomic RNA. Virology 188:102–113.
17. **Choo, Q.-L., G. Kuo, A. J. Weiner, L. R. Overby, D. W. Bradley, and M. Houghton.** 1989. Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. Science 244:359–362.
18. **Choo, Q.-L., K. H. Richman, J. H. Han, K. Berger, C. Lee, C. Dong, C. Gallegos, D. Coit, A. Medina-Selby, P. J. Barr, A. J. Weiner, D. W. Bradley, G. Kuo, and M. Houghton.** 1991. Genetic organization and diversity of the hepatitis C virus. Proc. Natl. Acad. Sci. USA 88:2451–2455.
19. **Ciechanover, A., J. A. DiGiuseppe, B. Bercovich, A. Orian, J. D. Richter, A. L. Schwartz, and G. M. Brodeur.** 1991. Degradation of nuclear oncoproteins by the ubiquitin system *in vitro*. Proc. Natl. Acad. Sci. USA 88:139–143.
20. **Collett, M. S.** 1992. Molecular genetics of pestiviruses. Comp. Immunol. Microbiol. Infect. Dis. 15:145–154.
21. **Collett, M. S., M. A. Wiskerchen, E. Welniak, and S. K. Belzer.** 1991. Bovine viral diarrhea virus genomic organization. Arch. Virol. 1991(Suppl. 3):19–27.
22. **Eckart, M. R., M. Selby, F. Masiarz, C. Lee, K. Berger, K. Crawford, C. Kuo, G. Kuo, M. Houghton, and Q.-L. Choo.** 1993. The hepatitis C virus encodes a serine protease involved in processing of the putative nonstructural proteins from the viral polyprotein precursor. Biochem. Biophys. Res. Commun. 192: 399–406.
23. **Failla, C., L. Tomei, and R. De Francesco.** 1994. Both NS3 and NS4A are required for proteolytic processing of hepatitis C virus nonstructural proteins. J. Virol. 68:3753–3760.
24. **Falgout, B., M. Pethel, Y.-M. Zhang, and C.-J. Lai.** 1991. Both nonstructural proteins NS2B and NS3 are required for the proteolytic processing of dengue virus nonstructural proteins. J. Virol. 65:2467–2475.
25. **Falkner, F. G., and B. Moss.** 1988. *Escherichia coli gpt* gene provides dominant selection for vaccinia virus open reading frame expression vectors. J. Virol. 62:1849–1854.
26. **Feinstone, S. M., H. J. Alter, H. P. Dienes, Y. Shimizu, H. Popper, D. Blackmore, D. Sly, W. T. London, and R. H. Purcell.** 1981. Non-A, non-B hepatitis in chimpanzees and marmosets. J. Infect. Dis. 144:588–598.
27. **Francki, R. I. B., C. M. Fauquet, D. L. Knudson, and F. Brown.** 1991. Classification and nomenclature of viruses: fifth report of the International Committee on Taxonomy of Viruses. Arch. Virol. 1991(Suppl. 2):223.
28. **Fuerst, T. R., E. G. Niles, F. W. Studier, and B. Moss.** 1986. Eukaryotic transient-expression system based on recombinant vaccinia virus that synthesizes bacteriophage T7 RNA polymerase. Proc. Natl. Acad. Sci. USA 83:8122–8126.
29. **Gorbalenya, A. E., A. P. Donchenko, E. V. Koonin, and V. M. Blinov.** 1989. N-terminal domains of putative helicases of flavi- and pestiviruses may be serine proteases. Nucleic Acids Res. 17:3889–3897.
30. **Grakoui, A., D. W. McCourt, C. Wychowski, S. M. Feinstone, and C. M. Rice.** 1993. Characterization of the hepatitis C virus-encoded serine proteinase: determination of proteinase-dependent polyprotein cleavage sites. J. Virol. 67:2832–2843.
31. **Grakoui, A., D. W. McCourt, C. Wychowski, S. M. Feinstone, and C. M. Rice.** 1993. A second hepatitis C virus-encoded proteinase. Proc. Natl. Acad. Sci. USA 90:10583–10587.
32. **Grakoui, A., C. Wychowski, C. Lin, S. M. Feinstone, and C. M. Rice.** 1993. Expression and identification of hepatitis C virus polyprotein cleavage products. J. Virol. 67:1385–1395.
33. **Hijikata, M., N. Kato, Y. Ootsuyama, M. Nakagawa, and K. Shimotohno.** 1991. Gene mapping of the putative structural region of the hepatitis C virus genome by *in vitro* processing analysis. Proc. Natl. Acad. Sci. USA 88:5547–5551.
34. **Hijikata, M., H. Mizushima, T. Akagi, S. Mori, N. Kakiuchi, N. Kato, T. Tanaka, K. Kimura, and K. Shimotohno.** 1993. Two distinct proteinase activities required for the processing of a putative nonstructural precursor protein of hepatitis C virus. J. Virol. 67:4665–4675.
35. **Hijikata, M., H. Mizushima, Y. Tanji, Y. Komoda, Y. Hirowatari, T. Akagi, N. Kato, K. Kimura, and K. Shimotohno.** 1993. Proteolytic processing and membrane association of putative nonstructural proteins of hepatitis C virus. Proc. Natl. Acad. Sci. USA 90:10773–10777.
36. **Houghton, M., A. Weiner, J. Han, G. Kuo, and Q.-L. Choo.** 1991. Molecular biology of the hepatitis C viruses: implications for diagnosis, development and control of viral disease. Hepatology 14:381–388.
37. **Hruby, D. E., L. A. Guarino, and J. R. Kates.** 1979. Vaccinia virus replication. I. Requirement for the host-cell nucleus. J. Virol. 29:705–715.
38. **Inchauspe, G., S. Zebedee, D.-H. Lee, M. Sugitani, M. Nasoff, and A. M. Prince.** 1991. Genomic structure of the human prototype strain H of hepatitis C virus: comparison with American and Japanese isolates. Proc. Natl. Acad. Sci. USA 88:10292–10296.
39. **Kato, N., M. Hijikata, Y. Ootsuyama, M. Nakagawa, S. Ohkoshi, T. Sugimura, and K. Shimotohno.** 1990. Molecular cloning of the human hepatitis C virus genome from Japanese patients with non-A, non-B hepatitis. Proc. Natl. Acad. Sci. USA 87:9524–9528.
40. **Kolykhalov, A. A., E. V. Agapov, and C. M. Rice.** 1994. Specificity of the hepatitis C virus NS3 serine protease: effects of substitutions at the 3/4A, 4A/4B, 4B/5A, and 5A/5B cleavage sites on polyprotein processing. J. Virol. 68:7525–7533.
41. **Kuo, G., Q.-L. Choo, H. J. Alter, G. L. Gitnick, A. G. Redeker, R. H. Purcell, T. Miyamura, J. L. Dienstag, M. J. Alter, C. E. Stevens, G. E. Tegtmeier, F. Bonino, M. Colombo, W.-S. Lee, C. Kuo, K. Berger, J. R. Shuster, L. R. Overby, D. W. Bradley, and M. Houghton.** 1989. An assay for circulating antibodies to a major etiologic virus of human non-A, non-B hepatitis. Science 244:362–364.
42. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) 227:680–685.
43. **Laskey, R. A., and A. D. Mills.** 1975. Quantitative film detection of ^{3}H and ^{14}C in polyacrylamide gels by fluorography. Eur. J. Biochem. 56:335–341.
44. **Lemm, J. A., T. Rümenapf, E. G. Strauss, J. H. Strauss, and C. M. Rice.** 1994. Polypeptide requirements for assembly of functional Sindbis virus replication complexes: a model for the temporal regulation of minus and plus-strand RNA synthesis. EMBO J. 13:2925–2934.
45. **Lin, C., S. M. Amberg, T. J. Chambers, and C. M. Rice.** 1993. Cleavage at a novel site in the NS4A region by the yellow fever virus NS2B-3 proteinase is a prerequisite for processing at the downstream 4A/4B signalase site. J. Virol. 67:2327–2335.
46. **Lin, C., T. J. Chambers, and C. M. Rice.** 1993. Mutagenesis of conserved residues at the yellow fever virus 3/4A and 4B/5 dibasic cleavage sites: effects on cleavage efficiency and polyprotein processing. Virology 192:596–604.
47. **Lin, C., B. D. Lindenbach, B. M. Prágai, D. W. McCourt, and C. M. Rice.** 1994. Processing of the hepatitis C virus E2-NS2 region: identification of p7 and two distinct E2-specific products with different C termini. J. Virol. 68:5063–5073.
48. **Lobigs, M.** 1992. Proteolytic processing of a Murray Valley encephalitis virus nonstructural polyprotein segment containing the viral proteinase: accumulation of a NS3-4A precursor which requires mature NS3 for efficient processing. J. Gen. Virol. 73:2305–2312.
49. **Mackett, M., and G. L. Smith.** 1986. Vaccinia virus expression vectors. J. Gen. Virol. 67:2067–2082.
50. **Manabe, S., I. Fuke, O. Tanishita, C. Kaji, Y. Gomi, S. Yoshida, C. Mori, A. Takamizawa, I. Yoshida, and H. Okayama.** 1994. Pro-

duction of nonstructural proteins of hepatitis C virus requires a putative viral protease encoded by NS3. Virology 198:636–644.
51. **Matsuura, Y., and T. Miyamura.** 1993. The molecular biology of hepatitis C virus. Semin. Virol. 4:297–304.
52. **Nestorowicz, A., T. J. Chambers, and C. M. Rice.** 1994. Mutagenesis of the yellow fever virus NS2A/2B cleavage site: effects on proteolytic processing, viral replication and evidence for alternative processing of the NS2A protein. Virology 199:114–123.
53. **Okamoto, H., M. Kojima, M. Sakamoto, H. Iizuka, S. Hadiwandowo, S. Suwignyo, Y. Miyakawa, and M. Mayumi.** 1994. The entire nucleotide sequence and classification of a hepatitis C virus isolate of a novel genotype from an Indonesian patient with chronic liver disease. J. Gen. Virol. 75:629–635.
54. **Okamoto, H., K. Kurai, S.-I. Okada, K. Yamamoto, H. Iizuka, T. Tanaka, S. Fukuda, F. Tsuda, and S. Mishiro.** 1992. Full-length sequence of a hepatitis C virus genome having poor homology to reported isolates: comparative study of four distinct genotypes. Virology 188:331–341.
55. **Okamoto, H., S. Okada, Y. Sugiyama, K. Kurai, H. Iizuka, A. Machida, Y. Miyakawa, and M. Mayumi.** 1991. Nucleotide sequence of the genomic RNA of hepatitis C virus isolated from a human carrier: comparison with reported isolates for conserved and divergent regions. J. Gen. Virol. 72:2697–2704.
56. **Preugschat, F., and J. H. Strauss.** 1991. Processing of nonstructural proteins NS4A and NS4B of dengue 2 virus *in vitro* and *in vivo*. Virology 185:689–697.
57. **Preugschat, F., C.-W. Yao, and J. H. Strauss.** 1990. In vitro processing of dengue virus type 2 nonstructural proteins NS2A, NS2B, and NS3. J. Virol. 64:4364–4374.
58. **Rice, C. M., R. Levis, J. H. Strauss, and H. V. Huang.** 1987. Production of infectious RNA transcripts from Sindbis virus cDNA clones: mapping of lethal mutations, rescue of a temperature-sensitive marker, and in vitro mutagenesis to generate defined mutants. J. Virol. 61:3809–3819.
59. **Rice, C. M., and J. H. Strauss.** 1982. Association of Sindbis virion glycoproteins and their precursors. J. Mol. Biol. 154:325–348.
60. **Sakamoto, M., Y. Akahane, F. Tsuda, T. Tanaka, D. G. Woodfield, and H. Okamoto.** 1994. Entire nucleotide sequence and characterization of a hepatitis C virus of genotype V/3a. J. Gen. Virol. 75:1761–1768.
61. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
62. **Schägger, H., and G. von Jagow.** 1987. Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range of 1 to 100 kDa. Anal. Biochem. 166:368–379.
63. **Shimotohno, K.** 1993. Hepatocellular carcinoma in Japan and its linkage to infection with hepatitis C virus. Semin. Virol. 4:305–312.
64. **Simmonds, P., E. C. Holmes, T.-A. Cha, S.-W. Chan, F. McOmish, B. Irvine, E. Beall, P. L. Yap, J. Kolberg, and M. S. Urdea.** 1993. Classification of hepatitis C virus into six major genotypes and a series of subtypes by phylogenetic analysis of the NS5 region. J. Gen. Virol. 74:2391–2399.
65. **Simmonds, P., D. B. Smith, F. McOmish, P. L. Yap, J. Kolberg, M. S. Urdea, and E. C. Holmes.** 1994. Identification of genotypes of hepatitis C virus by sequence comparisons in the core, E1 and NS5 regions. J. Gen. Virol. 75:1053–1061.
66. **Strauss, J. H., and E. G. Strauss.** 1994. The alphaviruses: gene expression, replication, and evolution. Microbiol. Rev. 58:491–562.
67. **Takamizawa, A., C. Mori, I. Fuke, S. Manabe, S. Murakami, J. Fujita, E. Onishi, T. Andoh, I. Yoshida, and H. Okayama.** 1991. Structure and organization of the hepatitis C virus genome isolated from human carriers. J. Virol. 65:1105–1113.
68. **Tanaka, T., N. Kato, M. Nakagawa, Y. Ootsuyama, M.-J. Cho, T. Nakazawa, M. Hijikata, Y. Ishimura, and K. Shimotohno.** 1992. Molecular cloning of hepatitis C virus genome from a single Japanese carrier: sequence variation within the same individual and among infected individuals. Virus Res. 23:39–53.
69. **Tomei, L., C. Failla, E. Santolini, R. De Francesco, and N. La Monica.** 1993. NS3 is a serine protease required for processing of hepatitis C virus polyprotein. J. Virol. 67:4017–4026.
70. **Wengler, G., G. Czaya, P. M. Färber, and J. H. Hegemann.** 1991. *In vitro* synthesis of West Nile virus proteins indicates that the amino-terminal segment of the NS3 protein contains the active centre of the protease which cleaves the viral polyprotein after multiple basic amino acids. J. Gen. Virol. 72:851–858.
71. **Wiskerchen, M., and M. S. Collett.** 1991. Pestivirus gene expression: protein p80 of bovine viral diarrhea virus is a proteinase involved in polyprotein processing. Virology 184:341–350.
72. **Zhang, L., P. M. Mohan, and R. Padmanabhan.** 1992. Processing and localization of dengue virus type 2 polyprotein precursor NS3-NS4A-NS4B-NS5. J. Virol. 66:7549–7554.

RESEARCH ARTICLE

Nucleotide Sequence of Yellow Fever Virus: Implications for Flavivirus Gene Expression and Evolution

Charles M. Rice, Edith M. Lenches, Sean R. Eddy
Se Jung Shin, Rebecca L. Sheets, James H. Strauss

Abstract. *The sequence of the entire RNA genome of the type flavivirus, yellow fever virus, has been obtained. Inspection of this sequence reveals a single long open reading frame of 10,233 nucleotides, which could encode a polypeptide of 3411 amino acids. The structural proteins are found within the amino-terminal 780 residues of this polyprotein; the remainder of the open reading frame consists of nonstructural viral polypeptides. This genome organization implies that mature viral proteins are produced by posttranslational cleavage of a polyprotein precursor and has implications for flavivirus RNA replication and for the evolutionary relation of this virus family to other RNA viruses.*

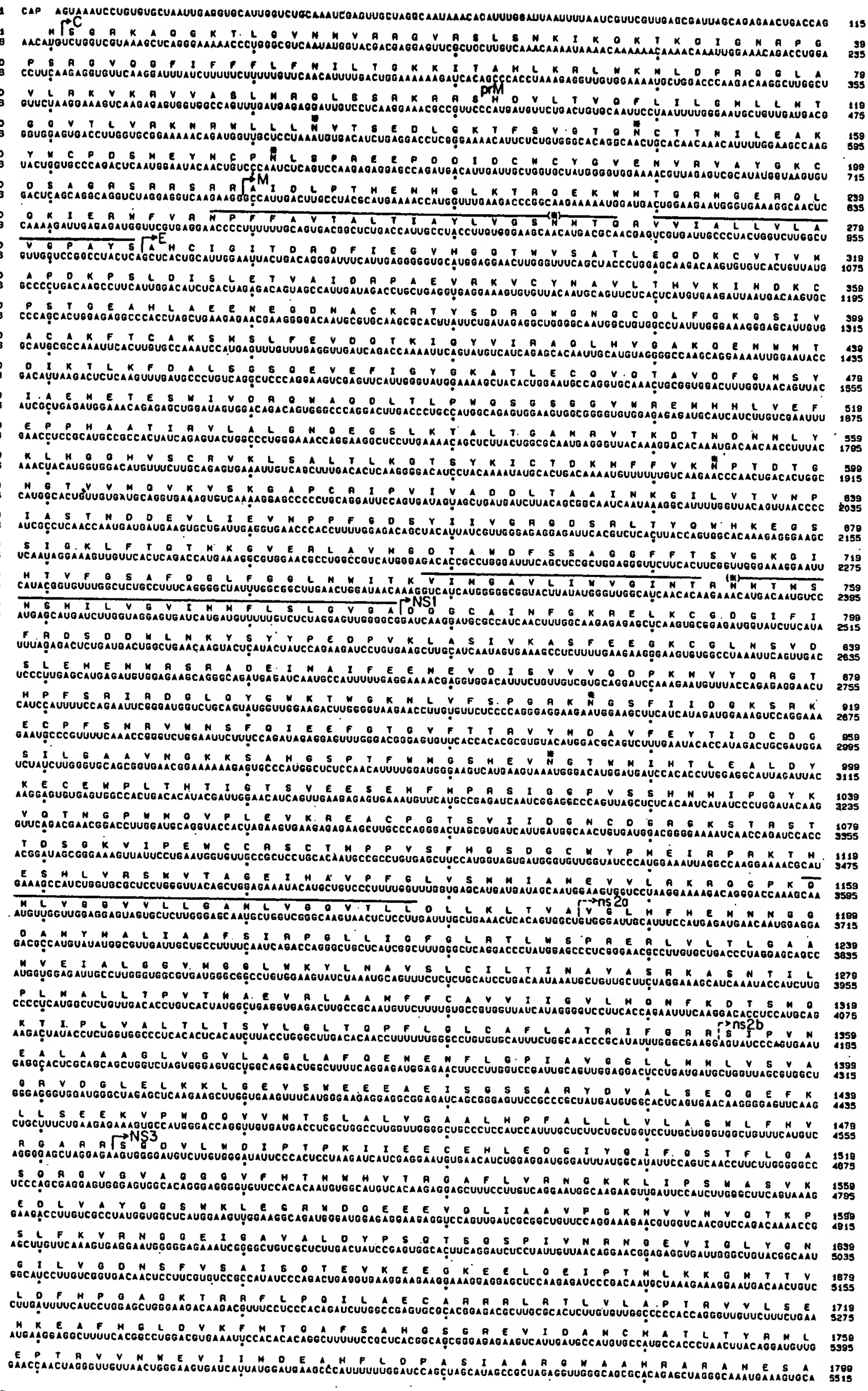
The *Flavivirus* genus, family Flaviviridae, consists of a group of some 70 closely related human or veterinary pathogens causing many serious illnesses, including dengue fever, Japanese encephalitis, St. Louis encephalitis, Murray Valley encephalitis, tick-borne encephalitis, and yellow fever (*1*). Most flaviviruses are transmitted to vertebrate hosts by blood-sucking arthropods, mosquitoes or ticks, although some evidently lack an arthropod vector (*2*). Arthropod-transmitted flaviviruses replicate in the arthropod host as well as the vertebrate host. Human flavivirus diseases have diverse and complex pathologies and different viruses exhibit marked tissue tropisms. Many are neurotropic, causing encephalitic symptoms; others, such as the dengue group, replicate preferentially in host macrophages, whereas yellow fever is usually viscerotropic.

The disease known as yellow fever has been recognized for several hundred years (*3*, *4*). Until the early 1900's recurrent epidemics occurred in the Caribbean area which caused great human suffering and had a profound influence on human activities in the area. From its focus in the Caribbean, epidemic yellow fever was spread by ship to ports as far north as Boston and as far east as England, where mortality rates in an epidemic could exceed 20 percent of those contracting the disease. Walter Reed and colleagues in pioneering studies in Cuba in 1900 demonstrated that yellow fever is transmitted by mosquitoes, and 2 years later showed that the disease agent is filterable (*5*). With the recognition that the mosquito *Aedes aegypti* is the vector for urban yellow fever, mosquito control measures rapidly led to the elimination of urban yellow fever. Subsequently, a safe and effective attenuated vaccine strain (17D) was developed by in vitro passage of the virulent Asibi strain in chicken embryo tissue (*6*). However, the virus persists in a sylvan cycle in the forests of South America and Africa, transmitted by numerous mosquito species including those of the genus *Haemagogus* in South America and of the genus *Aedes* in Africa. The vertebrate hosts in this cycle appear to be almost exclusively primates, demonstrating the limited natural host range of yellow fever. From the sylvan cycle periodic outbreaks in neighboring human populations have arisen on both continents. Furthermore, since *Aedes aegypti* is widespread in the world, a situation exacerbated by relaxation of mosquito abatement procedures in the Caribbean and elsewhere, the potential exists for future epidemics of urban yellow fever.

C. M. Rice, E. M. Lenches, and J. H. Strauss are members of the Division of Biology, California Institute of Technology, Pasadena 91125. S. R. Eddy and S. J. Shin are students at the California Institute of Technology and R. L. Sheets is doing graduate work in the Department of Cellular, Viral and Molecular Biology, University of Utah, University Medical Center, Salt Lake City 84132.

Previous studies have shown that flaviviruses contain single-stranded infectious RNA (thus defining them as plus-stranded RNA viruses in which the virion RNA serves as a messenger) encapsidated in a nucleocapsid possessing icosahedral symmetry and containing a single species of capsid protein [C, apparent mass of about 14 kilodaltons (kD)]. This in turn is surrounded by a lipid bilayer containing an envelope protein (E; about 50 to 60 kD) that is usually but not invariably glycosylated (*7*) and a second, nonglycosylated protein (M; about 8 kD) (*8*, *9*). How the envelope is obtained is unclear, as budding flaviviruses are seldom identified in electron microscopic studies, although maturation does appear to occur in association with intracellular membranes (*9*, *10*). Replication of flaviviruses in tissue culture is slow, with a long latent period, and only moderate titers of virus are produced. Host cell protein and RNA synthesis are shut off only poorly (vertebrate cells) or not at all (mosquito cells), making study of flavivirus replication and structure somewhat more difficult. Virus-specific protein synthesis appears to be associated with the rough endoplasmic reticulum, and RNA replication is localized in the perinuclear region (*11*). No subgenomic RNA has been detected in cells infected with flaviviruses, and it is believed that the genomic length RNA which is capped but not polyadenylated (*12*, *13*) is the only messenger RNA (mRNA) species (*9*, *12*, *14*). This mRNA is translated into the three structural proteins and several nonstructural proteins. Translation of the flavivirus genome in vitro produces polypeptides related to the structural proteins (*15*) which, in the presence of appropriate membrane fractions, can be processed efficiently to yield C and E (*16*). Peptide mapping of in vitro translation products as well as selective incorporation of *N*-formylmethionine suggest that initiation in vitro occurs only with the capsid protein. Alternatively, studies on the in vivo translation of flavivirus Kunjin have been based on the use of pactamycin or high salt inhibition of translation initiation (*17*) or ultraviolet inactivation of translation (*18*) in an attempt to map the genome order of flavivirus proteins on the assumption that there is just a single site for initiation of translation. These experiments have led Westaway and collaborators to suggest that multiple independent translation initiation sites are used within flavivirus RNA, a situation not typically found with other eukaryotic mRNA's (*19*).

We now present the complete nucleo-


C
prM
M
E
NS1
NS2a
NS2b
NS3

1800 T I L M T A T P P G T S D E F P H S N G E I E D V Q T D I P S E P W N T G H D W 1839
5516 ACAAUCUUGAUGACAGCCACACCGCCUGGGACUAGUGAUGAAUUUCCACAUUCAAAUGGUGAAAUAGAAGAUGUUCAAACGGACAUACCCAGUGAGCCCUGGAACACAGGGCAUGACUGG 5635

1840 I L A D K R P T A W F L P S I R A A N V M A A S L R K A G K S V V V L N R K T F 1879
5636 AUCCUAGCUGACAAAAGGCCCACGGCAUGGUUCCUUCCAUCCAUCAGAGCUGCAAAUGUCAUGGCUGCCUCUUUGCGUAAGGCUGGAAAGAGUGUGGUGGUCCUGAACAGGAAAACCUUU 5755

1880 E R E Y P T I K Q K K P D F I L A T D I A E M G A N L C V E R V L D C R T A F K 1919
5756 GAGAGAGAAUACCCCACGAUAAAGCAGAAGAAACCUGACUUUAUAUUGGCCACUGACAUAGCUGAAAUGGGAGCCAACCUUUGCGUGGAGCGAGUGCUGGAUUGCAGGACGGCUUUUAAG 5875

1920 P V L V D E G R K V A I K G P L R I S A S S A A Q R R G R I G R N P N R D G D S 1959
5876 CCUGUGCUUGUGGAUGAAGGGAGGAAGGUGGCAAUAAAAGGGCCACUUCGUAUCUCCGCAUCCUCUGCUGCUCAAAGGAGGGGGCGCAUUGGGAGAAAUCCCAACAGAGAUGGAGACUCA 5995

1960 Y Y Y S E P T S E N N A H H V C W L E A S M L L D N M E V R G G M V A P L Y G V 1999
5996 UACUACUAUUCUGAGCCUACAAGUGAAAAUAAUGCCCACCACGUCUGCUGGUUGGAGGCCUCAAUGCUCUUGGACAACAUGGAGGUGAGGGGUGGAAUGGUCGCCCCACUCUAUGGCGUU 6115

2000 E G T K T P V S P G E M R L R D D Q R K V F R E L V R N C D L P V W L S W Q V A 2039
6116 GAAGGAACUAAAACACCAGUUUCCCCUGGUGAAAUGAGACUGAGGGAUGACCAGAGGAAAGUCUUCAGAGAACUAGUGAGGAAUUGUGACCUGCCCGUUUGGCUUUCGUGGCAAGUGGCC 6235

2040 K A G L K T N D R K W C F E G P E E H E I L N D S G E T V K C R A P G G A K K P 2079
6236 AAGGCUGGUUUGAAGACGAAUGAUCGUAAGUGGUGUUUUGAAGGCCCUGAGGAACAUGAGAUCUUGAAUGACAGCGGUGAAACAGUGAAGUGCAGGGCUCCUGGAGGAGCAAAGAAGCCU 6355

→ns4a
2080 L R P R W C D E R V S S D Q S A L S E F I K F A E G R R ¦ G A A E V L V V L S E L 2119
6356 CUGCGCCCAAGGUGGUGUGAUGAAAGGGUGUCAUCUGACCAGAGUGCGCUGUCUGAAUUUAUUAAGUUUGCUGAAGGUAGGAGGGGAGCUGCUGAAGUGCUAGUUGUGCUGAGUGAACUC 6475

2120 P D F L A K K G G E A M D T I S V F L H S E E G S R A Y R N A L S M M P E A M T 2159
6476 CCUGAUUUCCUGGCUAAAAAAGGUGGAGAGGCAAUGGAUACCAUCAGUGUGUUCCUCCACUCUGAGGAAGGCUCUAGGGCUUACCGCAAUGCACUAUCAAUGAUGCCUGAGGCAAUGACA 6595

2160 I V M L F I L A G L L T S G M V I F F M S P K G I S R M S M A M G T M A G C G Y 2199
6596 AUAGUCAUGCUGUUUAUACUGGCUGGACUACUGACAUCGGGAAUGGUCAUCUUUUUCAUGUCUCCCAAAGGCAUCAGUAGAAUGUCUAUGGCGAUGGGCACAAUGGCCGGCUGUGGAUAU 6715

2200 L M F L G G V K P T H I S Y V M L I F F V L M V V V I P E P G Q Q R S I Q D N Q 2239
6716 CUCAUGUUCCUUGGAGGCGUCAAACCCACUCACAUCUCCUAUGUCAUGCUCAUAUUCUUUGUCCUGAUGGUGGUUGUGAUCCCCGAGCCAGGGCAACAAAGGUCCAUCCAAGACAACCAA 6835

2240 V A Y L I I G I L T L V S A V A A N E L G M L E K T K E D L F G K K N L I P S S 2279
6836 GUGGCAUACCUCAUUAUUGGCAUCCUGACGCUGGUUUCAGCGGUGGCAGCCAACGAGCUAGGCAUGCUGGAGAAAACCAAAGAGGACCUCUUUGGGAAGAAGAACUUAAUUCCAUCUAGU 6955

2280 A S P W S W P D L D L K P G A A W T V Y V G I V T M L S P M L H H W I K V E Y G 2319
6956 GCUUCACCCUGGAGUUGGCCGGAUCUUGACCUGAAGCCAGGAGCUGCCUGGACAGUGUACGUUGGCAUUGUUACAAUGCUCUCUCCAAUGUUGCACCACUGGAUCAAAGUCGAAUAUGGC 7075

2320 N L S L S G I A Q S A S V L S F M D K G I P F M K M N I S V I M L L V S G W N S 2359
7076 AACCUGUCUCUGUCUGGAAUAGCCCAGUCAGCCUCAGUCCUUUCUUUCAUGGACAAGGGGAUACCAUUCAUGAAGAUGAAUAUCUCGGUCAUAAUGCUGCUGGUCAGUGGCUGGAAUUCA 7195

→ns4b
2360 I T V M P L L C G I G C A M L H W S L I L P G I K A Q Q S K L A Q R R ¦ V F H G V 2399
7196 AUAACAGUGAUGCCUCUGCUCUGUGGCAUAGGGUGCGCCAUGCUCCACUGGUCUCUCAUUUUACCUGGAAUCAAAGCGCAGCAGUCAAAGCUUGCACAGAGAAGGGUGUUCCAUGGCGUU 7315

2400 A E N P V V D G N P T V D I E E A P E M P A L Y E K K L A L Y L L L A L S L A S 2439
7316 GCCGAGAACCCUGUGGUUGAUGGGAAUCCAACAGUUGACAUUGAGGAAGCUCCUGAAAUGCCUGCCCUUUAUGAGAAGAAACUGGCUCUAUAUCUCCUUCUUGCUCUCAGCCUAGCUUCU 7435

2440 V A M C R T P F S L A E G I V L A S A A L G P L I E G N T S L L W N G P M A V S 2479
7436 GUUGCCAUGUGCAGAACGCCCUUUUCAUUGGCUGAAGGCAUUGUCCUAGCAUCAGCUGCCUUAGGGCCGCUCAUAGAGGGAAACACCAGCCUUCUUUGGAAUGGACCCAUGGCUGUCUCC 7555

→NS5
2480 M T G V M R G N H Y A F V G V M Y N L W K M K T G R R | G S A N G K T L G E V W K 2519
7556 AUGACAGGAGUCAUGAGGGGGAAUCACUAUGCUUUUGUGGGAGUCAUGUACAAUCUAUGGAAGAUGAAAACUGGACGCCGGGGGAGCGCGAAUGGAAAAACUUUGGGUGAAGUCUGGAAG 7675

2520 R E L N L L D K R Q F E L Y K R T D I V E V D R D T A R R H L A E G K V D T G V 2559
7676 AGGGAACUGAAUCUGUUGGACAAGCGACAGUUUGAGUUGUAUAAAAGGACCGACAUUGUGGAGGUGGAUCGUGAUACGGCACGCAGGCAUUUGGCCGAAGGGAAGGUGGACACCGGGGUG 7795

2560 A V S R G T A K L R W F H E R G Y V K L E G R V I D L G C G R G G W C Y Y A A A 2599
7796 GCGGUCUCCAGGGGGACCGCAAAGUUAAGGUGGUUCCAUGAGCGUGGCUAUGUCAAGCUGGAAGGUAGGGUGAUUGACCUGGGGUGUGGCCGCGGAGGCUGGUGUUACUACGCUGCUGCG 7915

2600 Q K E V S G V K G F T L G R D G H E K P M N V Q S L G W N I I T F K D K T D I H 2639
7916 CAAAAGGAAGUGAGUGGGGUCAAAGGAUUUACUCUUGGAAGAGACGGCCAUGAGAAACCCAUGAAUGUGCAAAGUCUGGGAUGGAACAUCAUCACCUUCAAGGACAAAACUGAUAUCCAC 8035

2640 R L E P V K C D T L L C D I G E S S S S S V T E G E R T V R V L D T V E K W L A 2679
8036 CGCCUAGAACCAGUGAAAUGUGACACCCUUUUGUGUGACAUUGGAGAGUCAUCAUCGUCAUCGGUCACAGAGGGGGAAAGGACCGUGAGAGUUCUUGAUACUGUAGAAAAAUGGCUGGCU 8155

2680 C G V D N F C V K V L A P Y M P D V L E K L E L L Q R R F G G T V I R N P L S R 2719
8156 UGUGGGGUUGACAACUUCUGUGUGAAGGUGUUAGCUCCAUACAUGCCAGAUGUUCUCGAGAAACUGGAAUUGCUCCAAAGGAGGUUUGGCGGAACAGUGAUCAGGAACCCUCUCUCCAGG 8275

2720 N S T H E M Y Y V S G A R S N V T F T V N Q T S R L L M R R M R R P T G K V T L 2759
8276 AAUUCCACUCAUGAAAUGUACUACGUGUCUGGAGCCCGCAGCAAUGUCACAUUUACUGUGAACCAAACAUCCCGCCUCCUGAUGAGGAGAAUGAGGCGUCCAACUGGAAAAGUGACCCUG 8395

2760 E A D V I L P I G T R S V E T D K G P L D K E A I E E R V E R I K S E Y M T S W 2799
8396 GAGGCUGACGUCAUCCUCCCAAUUGGGACACGCAGUGUUGAGACAGACAAGGGACCCCUGGACAAAGAGGCCAUAGAAGAAAGGGUUGAGAGGAUAAAAUCUGAGUACAUGACCUCUUGG 8515

2800 F Y D N D N P Y R T W H Y C G S Y V T K T S G S A A S M V N G V I K I L T Y P W 2839
8516 UUUUAUGACAAUGACAACCCCUACAGGACCUGGCACUACUGUGGCUCCUAUGUCACAAAAACCUCAGGAAGUGCGGCGAGCAUGGUAAAUGGUGUUAUUAAAAUUCUGACAUAUCCAUGG 8635

2840 D R I E E V T R M A M T D T T P F G Q Q R V F K E K V D T R A K D P P A G T R K 2879
8636 GACAGGAUAGAGGAGGUCACAAGAAUGGCAAUGACUGACACAACCCCUUUUGGACAGCAAAGAGUGUUUAAAGAAAAAGUUGACACCAGAGCAAAGGAUCCACCAGCGGGAACUAGGAAG 8755

2880 I M K V V N R W L F R H L A R E K N P R L C T K E E F I A K V R S H A A I G A Y 2919
8756 AUCAUGAAAGUUGUCAACAGGUGGCUGUUCCGCCACCUGGCCAGAGAAAAGAACCCCAGACUGUGCACAAAGGAAGAAUUUAUUGCAAAAGUCCGAAGUCAUGCAGCCAUUGGAGCUUAC 8875

2920 L E E Q E Q W K T A N E A V Q D P K F W E L V D E E R K L H Q Q G R C R T C V Y 2959
8876 CUGGAAGAACAAGAACAGUGGAAGACUGCCAAUGAGGCUGUCCAAGACCCAAAGUUCUGGGAACUGGUGGAUGAAGAAAGGAAGCUGCACCAACAAGGCAGGUGUCGGACUUGUGUGUAC 8995

2960 N M M G K R E K K L S E F G K A K G S R A I W Y M W L G A R Y L E F E A L G F L 2999
8996 AACAUGAUGGGGAAAAGAGAGAAGAAGCUGUCAGAGUUUGGAAAGGCAAAGGGAAGCCGUGCCAUAUGGUAUAUGUGGCUGGGAGCGCGGUAUCUUGAGUUUGAGGCCCUGGGAUUCCUG 9115

3000 N E D H W A S R E N S G G G V E G I G L Q Y L G Y V I R D L A A M D G G G [F Y A 3039
9116 AAUGAGGACCAUUGGGCUUCCAGGGAAAACUCAGGAGGAGGAGUGGAAGGCAUUGGCUUACAAUACCUAGGAUAUGUGAUCAGAGACCUGGCUGCAAUGGAUGGUGGUGGAUUCUACGCG 9235

3040 D D T A G W D T R I T E A D L D D E Q E I L N Y M S P H H K K L A Q A V M E M T 3079
9236 GAUGACACCGCUGGAUGGGACACGCGCAUCACAGAGGCAGACCUUGAUGAUGAACAGGAGAUCUUGAACUACAUGAGCCCACAUCACAAAAAACUGGCACAAGCAGUGAUGGAAAUGACA 9355

3080 Y K N K V V K V L R P A P G G K A Y M D V I S R R D Q R G S G Q V V T Y A L N T 3119
9356 UACAAGAACAAAGUGGUGAAAGUGUUGAGACCAGCCCCAGGAGGGAAAGCCUACAUGGAUGUCAUAAGUCGACGAGACCAGAGAGGAUCCGGGCAGGUAGUGACUUAUGCUCUGAACACC 9475

3120 I T N L K V Q L I R M A E A E M V I H H Q H V Q D C D E S V L T R L E A W L T E 3159
9476 AUCACCAACUUGAAAGUCCAAUUGAUCAGAAUGGCAGAAGCAGAGAUGGUGAUACAUCACCAACAUGUUCAAGAUUGUGAUGAAUCAGUUCUGACCAGGCUGGAGGCAUGGCUCACUGAG 9595

3160 H G C D R L K R M A V S [G D D] C V V R P I D] D R F G L A L S H L N A M S K V R K 3199
9596 CACGGAUGUGACAGACUGAAGAGGAUGGCGGUGAGU[GGAGACGAC]UGUGUGGUCCGGCCCAUCGAUGACAGGUUCGGCCUGGCCCUGUCCCAUCUCAACGCCAUGUCCAAGGUUAGAAAG 9715

3200 D I S E W Q P S K G W N D W E N V P F C S H H F H E L Q L K D G R R I V V P C R 3239
9716 GACAUAUCUGAAUGGCAGCCAUCAAAAGGGUGGAAUGAUUGGGAGAAUGUGCCCUUCUGUUCCCACCACUUCCAUGAACUACAGCUGAAGGAUGGCAGGAGGAUUGUGGUGCCUUGCCGA 9835

3240 E Q D E L I G R G R V S P G N G W M I K E T A C L S K A Y A N M W S L M Y F H K 3279
9836 GAACAGGACGAGCUCAUUGGGAGAGGAAGGGUGUCUCCAGGAAACGGCUGGAUGAUCAAGGAAACAGCUUGCCUCAGCAAAGCCUAUGCCAACAUGUGGUCACUGAUGUAUUUUCACAAA 9955

3280 R D M R L L S L A V S S A V P T S W V P Q G R T T W S I H G K G E W M T T E D M 3319
9956 AGGGACAUGAGGCUACUGUCAUUGGCUGUUUCCUCAGCUGUUCCCACCUCAUGGGUUCCACAAGGACGCACAACAUGGUCGAUUCAUGGGAAAGGGGAGUGGAUGACCACGGAAGACAUG 10075

3320 L E V W N R V W I T N N P H M Q D K T M V K K W R D V P Y L T K R Q D K L C G S 3359
10076 CUUGAGGUGUGGAACAGAGUAUGGAUAACCAACAACCCACACAUGCAGGACAAGACAAUGGUGAAAAAAUGGAGAGAUGUCCCUUAUCUAACCAAGAGACAAGACAAGCUGUGCGGAUCA 10195

3360 L I G M T N R A T W A S H I H L V I H R I R T L I G Q E K Y T D Y L T V M D R Y 3399
10196 CUGAUUGGAAUGACCAAUAGGGCCACCUGGGCCUCCCACAUCCAUUUAGUCAUCCAUCGUAUCCGAACGCUGAUUGGACAGGAGAAAUACACUGACUACCUAACAGUCAUGGACAGGUAU 10315

3400 S V D A D L Q L G E L I 3411
10316 UCUGUGGAUGCUGACCUGCAACUGGGUGAGCUUAUC[UGA]AACACCAUC[UAA]CAGGA[AUAA]CCGGGAUACAAACCACGGGUGGAGAACCGGACUCCCCACAACCUGAAACCGGGAUAUAAA 10435

10436 CCACGGCUGGAGAACCGGGCUCCGCACUUAAAAUGAAACAGAAACCGGGAUAAAAACUACGGAUGGAGAACCGGACUCCACACAUUGAGACAGAAGAAGUUGUCAGCCCAGAACCCCACA 10555

10556 CGAGUUUUGCCACUGCUAAGCUGUGAGGCAGUGCAGGCUGGGACAGCCGACCUCCAGGUUGCGAAAAACCUGGUUUCUGGGACCUCCCACCCCAGAGUAAAAAGAACGGAGCCUCCGCUA 10675

10676 CCACCCUCCCACGUGGUGGUAGAAAGACGGGGUCUAGAGGUUAGAGGAGACCCUCCAGGGAACAAAUAGUGGGACCAUAUUGACGCCAGGGAAAGACCGGAGUGGUUCUCUGCUUUUCCU 10795

10796 CCAGAGGUCUGUGAGCACAGUUUGCUCAAGAAUAAGCAGACCUUUGGAUGACAAACACAAAACCACU 10862

tide sequence of the yellow fever genome determined from complementary DNA (cDNA) clones of the 17D vaccine strain. Together with recent NH_2-terminal sequence analysis of both structural (*20*) and some nonstructural yellow fever proteins, the amino acid sequences of the encoded proteins have been deduced and a preliminary picture of flavivirus gene organization and expression has begun to emerge.

Sequence of yellow fever RNA. The complete sequence of yellow fever RNA is shown in Fig. 1. The 5′- and 3′-terminal sequences presented were derived from several independent clones, are homologous to the 5′ and 3′ termini of West Nile flavivirus genomic RNA (*21*) (see below), and thus probably reflect the extreme ends of the yellow fever genome. Given these assumptions, the RNA genome is 10,862 nucleotides in length and has a mass of 3.75×10^6 daltons (expressed as the sodium form). Previous reports have shown that flavivirus genomic RNA contains a type 1 cap at the 5′ terminus but lacks a polyadenylate tract at the 3′ terminus (*12*, *13*). The base composition of the RNA is 27.3 percent A, 23.0 percent U, 28.4 percent G, and 21.3 percent C.

It is striking that the RNA contains an extremely long open reading frame, which spans virtually the entire length of the genome. This open reading frame, beginning from the first AUG triplet, is 10,233 nucleotides in length, terminating with a single opal codon (UGA), and could encode a polypeptide of 380,763 daltons, leaving 5′- and 3′-noncoding regions of 118 and 511 nucleotides, respectively. Examination of the remaining five possible reading frames (two in the virion RNA and three in the complementary RNA) reveals multiple stop codons in every case, with the longest possible other open reading frame being 804 nucleotides (in the complementary strand). Thus there is no reason to expect that any protein is translated from yellow fever RNA other than the polyprotein encoded by the long open reading frame shown in Fig. 1.

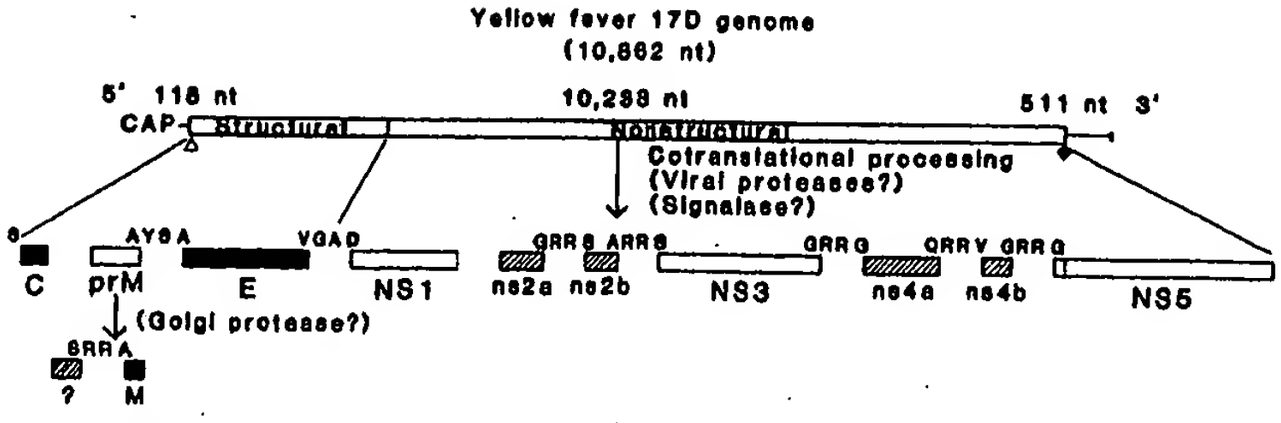


Fig. 2. Organization and processing of proteins encoded by the yellow fever genome. Untranslated regions are shown as single lines and the translated region as an open box. The open triangle is the initiation codon (AUG); the solid diamond the termination codon (UGA). The protein nomenclature is described in Table 1 and (*35*). The single letter amino acid code is used for sequences flanking assigned cleavage sites (solid lines). Two other potential cleavage sites are shown as dotted lines. Structural proteins, identified nonstructural proteins, and hypothesized nonstructural proteins (see text) are indicated by solid, open, and hatched boxes, respectively. Other potential cleavage sites have been found and are described in Table 1, footnote asterisk.

The structural proteins of yellow fever virus. The start points of the three yellow fever virus structural proteins (C, M, and E) have been positioned within the translated RNA sequence from NH_2-terminal amino acid sequences obtained for the structural proteins isolated from yellow fever virions (*20*) (Fig. 1). The capsid protein is the first protein found in the long open reading frame and begins one residue past the first methionine. Thus, in agreement with in vitro translation data from the flavivirus genomic RNA's of tick-borne encephalitis virus, West Nile virus, and Kunjin virus (*15*, *16*), the translation of the yellow fever genome initiates with the capsid protein, and the NH_2-terminal methionine is removed during maturation of the protein (*20*). The capsid protein may be released from the precursor polyprotein by cleavage at or just past a series of basic amino acids (Figs. 1 and 2). From this deduced amino acid sequence, the capsid protein is quite basic containing about 25 percent lysine and arginine distributed throughout the protein. The capsid protein of tick-borne encephalitis virus contains a similar proportion of basic amino acids (*22*). Since the capsid protein forms complexes with the RNA, its highly basic character probably acts to neutralize some of the RNA charges in such a compact structure.

Fig. 1 (preceding page and opposite page). Entire sequence of the genome of yellow fever virus. Yellow fever virus, 17D vaccine strain, was obtained from the American Type Culture Collection. This sample represents in vitro passage 234 of the line originated by Theiler and colleagues who started with the virulent Asibi strain (*6*). After plaque purification in Vero cells and amplification in BHK cells, the virus was grown in SW13 monolayers (*50*) and purified by polyethylene glycol precipitation, in glycerol-tartrate gradients. The purified virus was diluted with aqueous buffer and sedimented in the ultracentrifuge; the RNA was isolated by phenol extraction (*51*). Briefly, single-stranded cDNA was synthesized with avian myeloblastosis virus reverse transcriptase using degraded calf thymus DNA for priming (*47*). Second strand synthesis was carried out essentially as previously described (*52*). After methylation of the Eco RI sites with Eco RI methylase, phosphorylated Eco RI linkers were added with T4 DNA ligase. Following complete digestion with Eco RI, the double-stranded cDNA was sized on an agarose gel and selected size fractions were inserted into the Eco RI site of a plasmid vector derived from pBR322. Colonies containing yellow fever–specific inserts were selected by colony hybridization and were characterized by restriction mapping to obtain clones which represented most of the yellow fever genome. Clones containing the 3′ end of the genome were constructed by poly(A)-tailing (polyadenylation) the genomic RNA with *Escherichia coli* poly(A) polymerase followed by synthesis of double-stranded cDNA with an oligo(dT) primer. Addition of the poly(A) tract was relatively inefficient but after digestion of the double-stranded cDNA with Bgl I, 3′-terminal Bgl I fragments were selectively cloned with a plasmid vector derived from cloned yellow fever DNA (*51*). Clones containing the 5′ end of the genome were constructed by primer extension followed by oligo(dC) tailing with terminal deoxynucleotidyl transferase and oligo(dG) primed second strand synthesis. The entire sequence was obtained by chemical sequencing of both strands of the DNA (*53*). In addition, sequence was obtained throughout from at least two clones. Wherever the sequence differed between two clones (due presumably to heterogeneity in the RNA population or errors introduced during cloning), a third and occasionally a fourth clone was sequenced in this area, and the preferred nucleotide is reported here. Nucleotides are numbered from the 5′ terminus. Amino acids are numbered from the first methionine in the polyprotein sequence. The beginning of each protein is labeled (see Table 1 and text for nomenclature); tentative assignments are indicated by dashed arrows. Putative hydrophobic membrane-associated segments in the structural region are overlined. Potential *N*-linked glycosylation sites are denoted by an asterisk. The region of NS5 homologous to other RNA viruses (see text) is enclosed by brackets and the conserved Gly-Asp-Asp sequence is boxed. Repeated nucleotide sequences are underlined. Closely spaced in phase stop codons that terminate the long open reading frame are boxed. The single letter abbreviations for the amino acid residues are: A, alanine; C, cysteine; D, aspartic acid; E, glutamic acid; F, phenylalanine; G, glycine; H, histidine; I, isoleucine; K, lysine; L, leucine; M, methionine; N, asparagine, P, proline; Q, glutamine; R, arginine; S, serine; T, threonine; V, valine; W, tryptophan; Y, tyrosine.

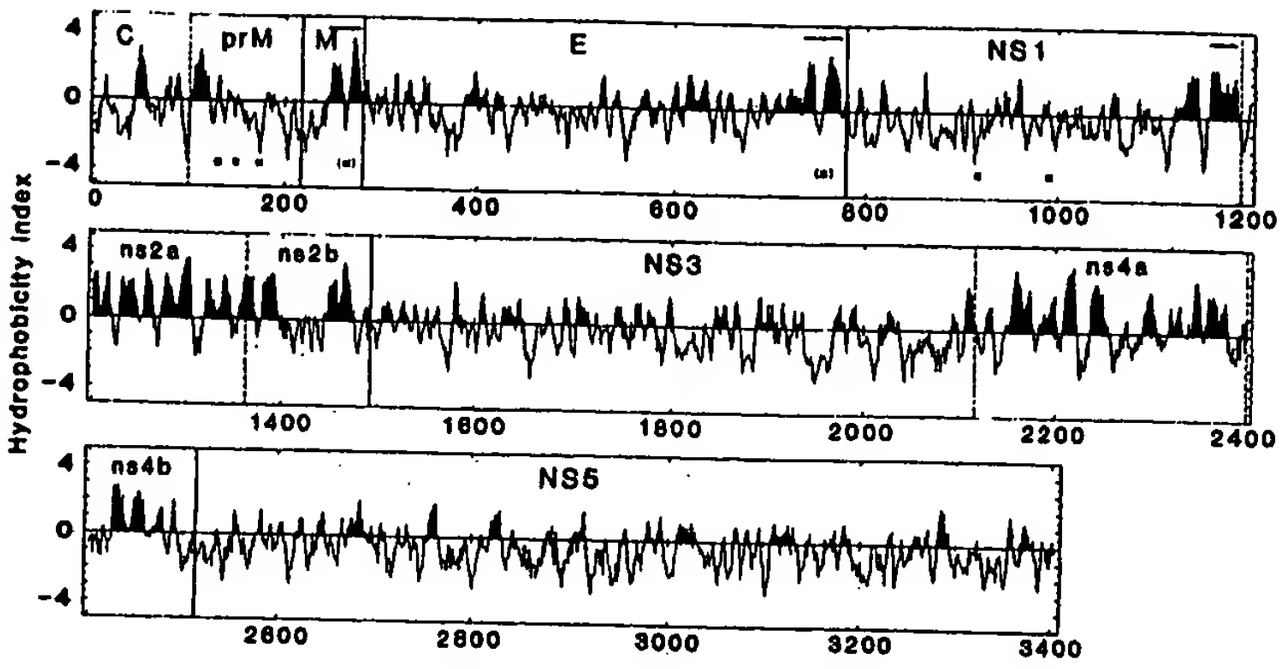


Fig. 3. Hydrophobicity plot of the yellow fever polyprotein sequence. The program of Kyte and Doolittle (*54*) with a search length of seven amino acids was used. Cleavage sites localized by NH_2-terminal protein sequence are indicated by solid vertical lines; putative cleavage sites are indicated by dotted vertical lines. The protein nomenclature is described in Table 1 and (*35*). The degree of hydrophobicity increases with distance above the horizontal line; hydrophilicity increases with distance below the horizontal line. Potential *N*-linked glycosylation sites are denoted by asterisks and putative membrane-associated anchors are indicated by solid bars.

There is also a hydrophobic stretch of 16 uncharged amino acids beginning with residue 42 from the NH_2 terminus (see Fig. 3), which is conserved among flaviviruses (*23*) and may be involved in protein-protein or specific protein-RNA interactions (or both) which assemble the nucleocapsid and lead to acquisition of the lipoprotein envelope by the capsid.

The start point of the virion M protein is also shown in Fig. 1. This protein contains a charged NH_2-terminal domain and two long uncharged stretches at its COOH terminus; these two stretches are separated by a single basic residue (Figs. 1 and 3) and could act as membrane spanning anchors similar to those observed in many virus envelope proteins. Protein M has not been identified in infected cells and is postulated to be derived from a precursor glycoprotein which we call prM (Table 1), which is also called by others GP23, GP19, or NV2 (*8, 24*). The sequence data support this hypothesis. A possible start point of prM, as deduced by limited homology with the NH_2-terminal sequence of the flavivirus St. Louis encephalitis NV2 (*20*) and homology in this region with Murray Valley encephalitis virus (*23*), follows the capsid protein; prM may begin with an uncharged stretch of amino acids which could function as an NH_2-terminal signal sequence for its cotranslational insertion into the endoplasmic reticulum (Fig. 3). After this hydrophobic domain, which may or may not be removed by signalase, the prM sequence contains three possible glycosylation sites of the type Asn-X-Ser/Thr. The NH_2 terminus of M (*20*) follows the sequence Arg-Ser-Arg-Arg in prM, indicating that the cleavage to produce M may be effected by the same enzyme that cleaves a number of viral envelope precursors at the sequence Arg-X-Arg/Lys-Arg and that has been postulated to be a host protease localized in the Golgi apparatus or Golgi-derived vesicles (*25*), perhaps similar to the cathepsins (*26*). As a result of this cleavage, which apparently occurs late during virus maturation and release, an 11.4-kD (not including carbohydrate) glycopeptide would be removed leaving the nonglycosylated M protein embedded in the virion membrane. Trace quantities of small virus-specific glycoproteins have been detected in cytoplasmic extracts (*27, 28, 29*), but whether the glycopeptide fragment remains cell-associated and is rapidly degraded or is released into the extracellular medium is unknown.

The E protein follows M. The NH_2 terminus of E is charged, and the more hydrophobic COOH-terminal domain of M (or its precursor, prM) may function as the signal sequence for the translocation of E across the rough endoplasmic reticulum. The protein E contains two sites of the form Asn-X-Ser/Thr which could serve as carbohydrate attachment sites, and both glycosylated and nonglycosylated forms have been detected in infected cells (*7, 27, 30*). The COOH-terminal domain of E contains uncharged stretches that could serve as a transmembrane anchor. Cleavage between M and E occurs after a Ser residue, and could be catalyzed by a host protease such as signalase. Since the COOH terminus of the mature M protein has not been determined, a small peptide, analogous to the 6 kD protein of alphaviruses (*25, 31*) could be produced during maturation of M and E. However, the apparent size of the M protein agrees well with the predicted molecular weight if cleavage occurs after the Ser at position 285.

This model for translation and processing of structural proteins and the features mentioned above predict that most of the E protein and some of the M protein should be exposed on the mature virion surface, and therefore sensitive to digestion by appropriate proteases. Protease digestion of purified tick-borne encephalitis virus (*32*) and also yellow fever virus (*29*) support this hypothesis. Thus, the M protein (or prM) of flaviviruses is an integral membrane protein and may interact specifically with both the E protein as well as the capsid protein-RNA complex during virus assembly.

The nonstructural proteins. In addition to prM, at least four and as many as 12 nonstructural proteins have been described in flavivirus-infected cells (*9, 28, 33, 34*). Some or all of these proteins must be active in the replication of the viral RNA. The start points of the three largest nonstructural proteins (NV3, NV4, and NV5 by the old nomenclature) (*35*) have been located by NH_2-terminal amino acid sequence analysis (*36*). As previously suggested by peptide mapping of the corresponding nonstructural proteins from other flaviviruses (*9, 15, 34*), the sequence data show that these proteins map to nonoverlapping segments in the yellow fever virus nonstructural region (Figs. 1 and 2).

In an attempt to simplify the description of flavivirus encoded nonstructural polypeptides, in particular the smaller proteins, we suggest a modified nomenclature (*35*) (Table 1) based on the linear order of these proteins in the yellow fever virus genome to complement designations based on their apparent molecular weights (*37*). In taking this approach we assume that members of Flaviviridae will have similar genome organization and express homologous proteins from homologous regions of their genomes. This assumption has been partially verified by an extensive sequence comparison of yellow fever virus with another member of the flavivirus genus, Murray Valley encephalitis virus (*23*).

Several features of the yellow fever virus nonstructural region are apparent from the localization of NS1, NS3, and NS5 (formerly NV3, NV4, and NV5). First, NS1 immediately follows the putative transmembrane segment of the E

protein. It should be noted that NS1 is glycosylated (*27*), and monoclonal antibodies against NS1 are capable of mediating complement-dependent lysis of yellow fever virus–infected cells, suggesting its presence at the plasma membrane (*38*). Thus, the COOH-terminal uncharged hydrophobic sequence of E could function as a signal sequence for translocation of NS1 across the endoplasmic reticulum. NS1 contains two sites of the type Asn-X-Ser/Thr which could serve as glycosylation sites. The probable COOH terminus of NS1 from estimates of molecular weight could contain a hydrophobic sequence for anchoring the protein in the membrane (Fig. 3). Thus the three glycoproteins of yellow fever virus, prM, E, and NS1, are adjacent to one another in the genome and are possibly inserted into the membrane one after another during synthesis. The sequence data support the hypothesis that each has the usual membrane protein topology of an NH_2 terminus outside and a COOH-terminal hydrophobic anchor. However, additional experiments are required to rigorously establish their orientation with respect to the lipid bilayer and exact COOH termini. The function of NS1 is unknown, but it could be involved in virus assembly rather than RNA replication. In this regard, it is of interest that NS1 has been shown to be the soluble complement-fixing antigen for dengue 2 (*28*) and suggestive evidence exists for a comparable role of NS1 in yellow fever virus infection (*8*, *27*). Thus, this protein may exist in alternative membrane-associated and soluble forms, perhaps because of the presence or absence of the COOH-terminal hydrophobic domain.

NS3 begins at residue 1485 in the polyprotein sequence and is produced by cleavage at the site Gly-Ala-Arg-Arg ↓ Ser; the NH_2-terminus of NS5 has been tentatively identified as residue 2507 after cleavage at Thr-Gly-Arg-Arg ↓ Gly. Since no host proteases with this specificity (which are active in the cytosol) have been characterized and animal viruses often encode proteases active in the processing of their cytoplasmic polyprotein precursors, yellow fever virus may encode a protease that cleaves after two Arg residues (or two basic residues) surrounded by amino acids with short side chains, often Gly (Table 1 and footnote asterisk).

These assignments leave two regions in the polyprotein for which polypeptide products have not yet been identified. Assuming that other nonstructural proteins will be produced from these regions by the same protease responsible for NH_2-terminal cleavage of NS3 and NS5, we have scanned the remaining sequences for additional cleavage sites. Estimates of molecular weight (*27*) have positioned the COOH terminus of NS1 near residue 1187. The next potential cleavage sequence, Gly-Arg-Arg ↓ Ser, at residue 1355 would produce two small nonstructural polypeptides of approximately 18 kD (ns2a) and 14 kD (ns2b) located between NS1 and NS3 (Fig. 2 and Table 1). Both of these polypeptides would be extremely hydrophobic (Fig. 3) with ns2b containing a short internal charged domain. The putative cleavage at the sequence Glu-Gly-Arg-Arg ↓ Gly (residue 2108) would produce a polypeptide whose calculated mass agrees well with the observed size of NS3 on polyacrylamide gels (*27*, *29*). Between this site and the NH_2 terminus of NS5 a single potential cleavage site (Ala-Gln-Arg-Arg ↓ Val) is found preceding residue 2395. Cleavage here would result in two methionine-rich, hydrophobic polypeptides of 31 kD (ns4a) and 12 kD (ns4b) (see Figs. 2 and 3 and Table 1). Polypeptides of these approximate sizes (10, 14, 18, and 30 kD) do exist in yellow fever–infected cells, but definitive mapping of these polypeptides as well as other minor species await additional NH_2-terminal sequence data. Similarly in the absence of COOH-terminal sequence data we cannot be sure of the exact terminal residues. Some heterogeneity in flavivirus polypeptides may result from variable exopeptidase digestion of the COOH-terminal residues or alternative internal cleavages. The predicted size of NS5, if the protein encompasses the remainder of the open reading frame, agrees well with its observed size (*27*).

Implications for flavivirus replication. It has been suggested that flavivirus RNA is translated by multiple internal initiation events (*17*, *18*) which would make flaviviruses atypical among eukaryotic viruses and eukaryotic genes. The presence of a single long open reading frame in yellow fever virus RNA, the

Table 1. Flavivirus polypeptides.

Protein nomenclature (*35*)		NH_2-terminal cleavage site*	M_r†	$M_{pred.}$‡	Glycosylated?	Comments
Proposed	Old					
			Structural region			
C	V2 (NV1½)	M ↓ S	13,000 to 16,000	11,320	No	Nucleocapsid protein
prM	(NV2) (NV2½)	?	19,000 to 23,000	20,925	Yes	Precursor to M
M	V1	SRR ↓ A	8,000 to 8,500	8,526	No	Virion envelope protein
E	V3	AYS ↓ A	51,000 to 60,000	53,712	Both forms§	Major virion envelope protein
			Nonstructural region			
NS1	NV3	VGA ↓ D	44,000 to 49,000	45,869	Yes	Soluble complement-fixing antigen
ns2a	(NV2½) (NV2)	(TVA ↓ V)	16,000 to 21,000	18,086	No	Hydrophobic; function unknown
ns2b	(NV1½)	(GRR ↓ S)	12,000 to 15,000	13,823	No	Hydrophobic; function unknown
NS3	NV4	ARR ↓ S	67,000 to 76,000	69,319	No	Replicase component ?
ns4a	(NVX) (NV2½)	(GRR ↓ G)	24,000 to 32,000	31,196	No	Hydrophobic; function unknown
ns4b	(NV1)	(QRR ↓ V)	10,000 to 11,000	12,159	No	Hydrophobic; function unknown
NS5	NV5	GRR ↓ G	91,000 to 98,000	104,079	No	Replicase component ?

*Cleavage sites predicted by NH_2-terminal protein sequence data (*20*, *36*). Tentative sites (indicated by parenthesis) are based on homology with confirmed cleavage sites and the sizes of yellow fever–specific polypeptides observed in infected cells (*27*, *29*). Alternative cleavage sites in the nonstructural region occur after residue 1946 (Gln-Arg-Arg ↓ Gly), residue 2548 (Ala-Arg-Arg ↓ His), residue 2707 (Gln-Arg-Arg ↓ Phe), and residue 3104 (Ser-Arg-Arg ↓ Asp). †Range of flavivirus protein sizes estimated from acrylamide gel electrophoresis. Some of these proteins have not yet been identified for all flaviviruses thus far examined [for comparative analyses see (*33*, *34*)]. In particular, definitive comparisons between NV3, NV2, NV2½, NVX, and NV1½ are difficult because of the complexity of flavivirus protein patterns in the 8,000 $< M_r <$ 45,000 size range. Alternative pathways of posttranslational cleavage may be used by different viruses for the production of the smaller nonstructural polypeptides. However, given the relatively consistent pattern of structural and larger nonstructural proteins, it seems likely that the small nonstructural proteins are also conserved, but their apparent migration on acrylamide gels and labeling efficiency are influenced by differences in amino acid composition. [For more complete discussion of this subject see (*49*).] ‡Polypeptide molecular weights calculated according to the cleavage sites shown in Fig. 1, with the C-prM cleavage site between residues 101 and 102. §Both glycosylated and nonglycosylated forms of E have been identified for yellow fever virus (*27*) and Kunjin virus (*7*, *30*).

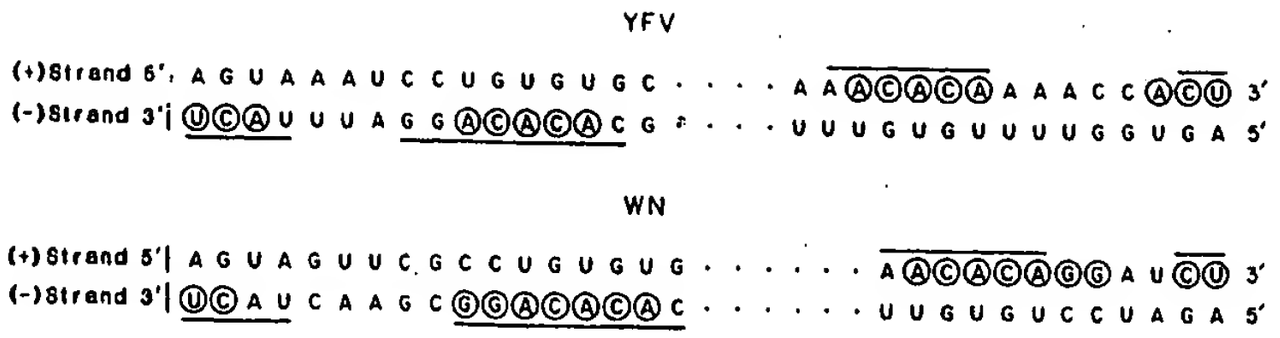


Fig. 4. Nucleotide homology between yellow fever virus (YFV) and West Nile virus (WN) [WN data are from (*21*)] at the 3′ termini of the genomic (+) strand and complementary (−) strand RNA's. Nucleotide identities in the 3′-terminal sequences of (+) and (−) strands are circled; those which are homologous between yellow fever and West Nile RNA's are underlined [(−) strand] or overlined [(+) strand].

fact that the final proteins found do not initiate with methionine but appear to arise from a consistent set of proteolytic cleavages, the gene order deduced from the pactamycin runoff experiments of Westaway (*17*), the in vitro translation data (*15*, *16*), and recent evidence for polyprotein precursors (*39*) all support the view that translation of the flavivirus genome in vivo initiates with the capsid protein near the 5′ end of the genome and proceeds sequentially through the genome to produce one precursor polyprotein.

Cleavage of this precursor is rapid and occurs during translation so that the precursor is not seen in its entirety. The location and frequency of characteristic cleavage sites in this precursor suggest that processing involves both virus encoded and cellular organelle bound proteases. Although internal translation initiation cannot be formally excluded, the 5′ terminal location of the structural genes and the 3′ terminal replicase genes implies that the relative amounts of structural and nonstructural gene products could also be regulated by premature termination as well as by nonuniform rates of translation (*40*) or differential stability of the final products. A potential secondary structure in yellow fever RNA just past the structural protein genes could possibly be active in the former mechanism. It is unclear why gene mapping experiments with ultraviolet light to inactivate translation (*18*) or high salt to synchronize initiation of translation (*17*) suggest multiple independent sites of initiation and do not allow prediction of the correct gene order. Possible explanations are that ribosomes might have slow transit velocities in some areas, due to RNA secondary structures or the presence of rare codons (*40*), or that it might be necessary to translate a functional protease to produce the final products.

Several features potentially important in RNA replication or packaging (or both) can be identified in the genomic sequence. First, the extreme 5′- and 3′-terminal sequences are homologous to those found for another flavivirus, West Nile virus (*21*) (Fig. 4), and the complement of the 5′-terminal sequence [equivalent to the 3′ terminus of the (−) strand] is related to the 3′-terminal sequence of the (+) strand. This suggests that the viral replicase may have similar recognition sites for (+) and (−) strand synthesis. In addition, a stable secondary structure (ΔG = −40 to 45 kcal) can be formed from the 3′-terminal 87 nucleotides of the yellow fever genomic RNA (Fig. 5). This may be involved in RNA replication as well as encapsidation, and if conserved among flaviviruses could explain the observation that many flavivirus RNA's are poor substrates for 3′-terminal enzymatic modification including ligation and addition of poly(A) (polyadenylate). Similar transfer RNA–like secondary structures and conserved sequences have been identified at the 3′ end of many plant viral RNA's (*41*); in addition to serving as substrates for aminoacylation both in vivo and in vitro (*42*) they are important for initiation of (−) strand RNA synthesis (*43*). Last, the 3′-untranslated region contains a set of three closely spaced repeated sequences (underlined in Fig. 1) (located between nucleotides 10,374 and 10,520) each approximately 40 nucleotides long with an average of six changes between them in pairwise comparisons. The significance of these repeats in flavivirus replication is unknown.

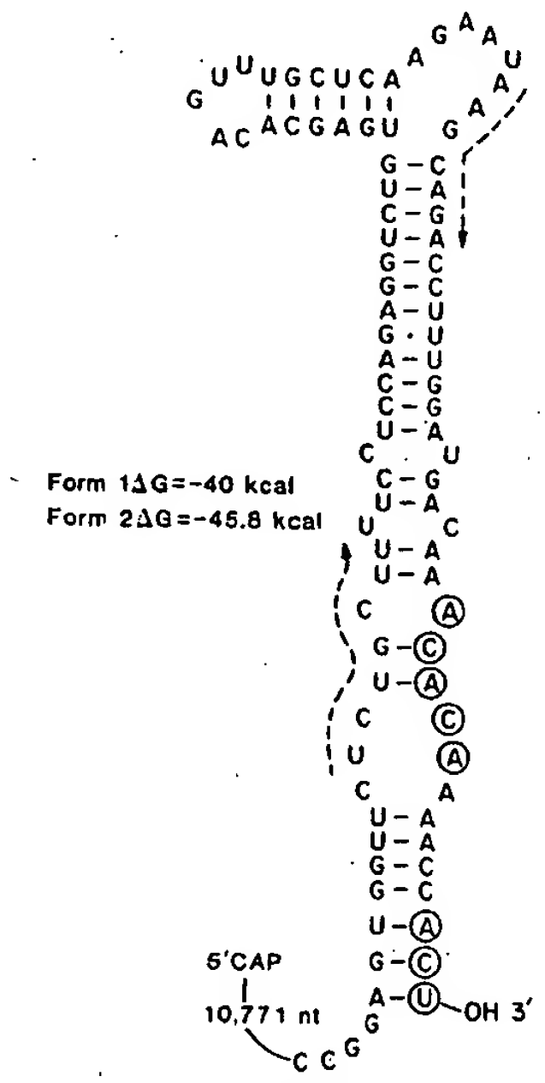


Fig. 5. Possible secondary structures at the 3′ terminus of yellow fever virus genomic RNA. Circled nucleotides are shared with the 3′ terminus of the yellow fever (−) strand (see Fig. 4). ΔG values were calculated according to Tinoco *et al.* (*55*). A more stable conformation than the one shown (form 1) can be formed if the two overlined sequences are base paired (form 2).

Evolution of flaviviruses. It is becoming clear that the flaviviruses deserve their recent reclassification as a family separate from the alphaviruses. Although the mature virions are morphologically similar to alphaviruses in that they have a single-stranded RNA (+) sense genome encapsidated in an icosahedral nucleocapsid and surrounded by a lipid bilayer containing virus-specified polypeptides, they differ markedly in genome organization and replication strategy (*44*). The location of the genes encoding the structural proteins at the 5′ end of the genome, the single long reading frame, and the lack of a subgenomic message are all characteristics shared with picornaviruses rather than togaviruses.

In order to understand the evolutionary role of flaviviruses and their relation to other RNA viruses we have searched for homologies within the putative polymerase genes of various plant and animal viruses. Significant homologies have been found between alphaviruses and plant viruses (*45*) and less extensive homologies between picornaviruses and alphaviruses (*46*). Kamer and Argos (*46*) have aligned the polymerase gene of poliovirus with those of several viruses including alfalfa mosaic virus, bromegrass mosaic virus, tobacco mosaic virus, Sindbis virus, foot and mouth disease virus, encephalomyocarditis virus, and cowpea mosaic virus. The amino acid sequence of yellow fever virus NS5 between residues 3037 and 3181 can also be aligned with this collection of diverse RNA viruses (Fig. 1). These homologous regions are convincing but short and probably represent conserved functional domains for particular RNA-dependent polymerase functions. It is interesting to

speculate on the origin of this diverse group of viruses. Whether they arose from one or a few protoviruses (perhaps insect viruses) and have radiated to their current divergent hosts or whether the viruses have repeatedly cannibalized their hosts, obtaining their replicases from eukaryotic cellular functions cannot be resolved at present. However, one possible measure of host adaptation or origin of viral genes from host functions is the CG doublet frequency in the RNA. Insects, insect viruses, and alphaviruses (insect-borne with vertebrate hosts) have the expected CG doublet frequency predicted from their base compositions (*47*), whereas vertebrate DNA (*48*), viruses with exclusively vertebrate hosts, and yellow fever virus have low CG doublet frequencies (2.4 percent CG found in yellow fever compared to 6.1 percent predicted from the base composition). Given the rapid evolution of RNA genomes, it is unlikely that this difference applies directly to the question of evolutionary origin of alphaviruses and flaviviruses but rather reflects alternative strategies of adaptation to their arthropod and vertebrate hosts in ways which are not currently understood.

Comparative studies with other flaviviruses should help to define areas of commonality of function in the nonstructural proteins, to localize biologically important antigenic epitopes on the structural polypeptides (and NS1) and to ascertain whether certain features of the yellow fever sequence (like the putative secondary structure at the extreme 3′ terminus and repeated nucleotide sequences) are functionally significant landmarks conserved among flaviviruses. In addition, the construction of cDNA clones designed for expression of functional virus gene products or production of infectious virus should provide useful new approaches for studying flavivirus molecular biology and pathogenesis as well as for development of flavivirus vaccines.

References and Notes

1. R. E. Shope, in *The Togaviruses*, R. W. Schlesinger, Ed. (Academic Press, New York, 1980), pp. 47–82.
2. R. W. Chamberlain, *ibid.*, pp. 175–227.
3. G. K. Strode, Ed., *Yellow Fever* (McGraw-Hill, New York, 1951).
4. W. G. Downs, *Yale J. Biol. Med.* 55, 179 (1982).
5. W. Reed, *Med. Rec.* 60, 201 (1901); _____ and J. Carroll, *Am Med.* 3, 301 (1902); C. Norman, *Science* 223, 1370 (1984).
6. M. Theiler and H. H. Smith, *J. Exp. Med.* 65, 767 (1937); *ibid.*, p. 787.
7. P. J. Wright, *J. Gen. Virol.* 59, 29 (1982).
8. P. K. Russell, W. E. Brandt, J. M. Dalrymple, in *The Togavirus*, R. W. Schlesinger, Ed. (Academic Press, New York, 1980), pp. 503–529.
9. E. G. Westaway, *ibid.*, pp. 531–581.
10. R. W. Boulton and E. G. Westaway, *Virology* 69, 416 (1976); F. A. Murphy, in *The Togaviruses*, R. W. Schlesinger, Ed. (Academic Press, New York, 1980), pp. 241–316.
11. M. L. Ng, J. S. Pedersen, B. H. Toh, E. G. Westaway, *Arch. Virol.* 78, 177 (1983).
12. G. Wengler, G. Wengler, H. J. Gross, *Virology* 89, 423 (1978).
13. G. R. Cleaves and D. T. Dubin, *ibid.* 96, 159 (1979); V. Deubel *et al.*, *Ann. Virol. (Inst. Pasteur)* 134E, 581 (1983).
14. R. W. Boulton and E. G. Westaway, *Arch. Virol.* 55, 201 (1977); C. W. Naeve and D. W. Trent, *J. Virol.* 25, 535 (1978).
15. Y. V. Svitkin *et al.*, *FEBS Lett.* 96, 211 (1978); G. Wengler, M. Beato, G. Wengler, *Virology* 96, 516 (1979); Y. V. Svitkin *et al.*, *Ibid.* 110, 26 (1981); R. P. Monckton and E. G. Westaway, *J. Gen. Virol.* 63, 227 (1982).
16. Y. V. Sitkin *et al.*, *Virology* 135, 536 (1984).
17. E. G. Westaway, *ibid.* 80, 320 (1977).
18. _____, G. Speight, L. Endo, *Virus Res.* 1, 333 (1984).
19. M. Kozak, *Microbiol. Rev.* 47, 1 (1983); we note a possible exception in the case of infectious pancreatic necrosis A RNA [P. P. C. Mertens and P. Dobos, *Nature (London)* 297, 243 (1982)].
20. J. R. Bell *et al.*, *Virology* 143, 224 (1985).
21. G. Wengler and G. Wengler, *ibid.* 113, 544 (1981).
22. U. Boege, F. X. Heinz, G. Wengler, C. Kunz, *ibid.* 126, 651 (1983).
23. L. Dalgarno, D. Trent, J. H. Strauss, C. M. Rice, in preparation.
24. D. Shapiro, W. E. Brandt, P. K. Russell, *Virology* 50, 906 (1972).
25. H. Garoff, A. M. Frischauf, K. Simons, H. Lehrach, H. Delius, *Nature (London)* 288, 235 (1980); C. M. Rice and J. H. Strauss, *Proc. Natl. Acad. Sci. U.S.A.* 78, 2062 (1981).
26. K. Takio, T. Towatari, N. Katunuma, D. C. Teller, K. Titani, *Proc. Natl. Acad. Sci. U.S.A.* 80, 3666 (1983).
27. J. J. Schlesinger, M. W. Brandriss, T. P. Monath, *Virology* 125, 8 (1983).
28. G. W. Smith and P. J. Wright, *J. Gen. Virol.* 66, 559 (1985).
29. J. Pata and C. M. Rice, unpublished data.
30. P. J. Wright, H. M. Warr, E. G. Westaway, *Virology* 109, 418 (1981).
31. W. J. Welch and B. M. Sefton, *J. Virol.* 29, 1186 (1979).
32. F. X. Heinz and C. Kunz, *Arch. Virol.* 60, 207 (1979); F. Heinz, personal communication.
33. E. G. Westaway, *Virology* 51, 454 (1973); _____, J. H. McKimm, L. G. McLeod, *Arch. Virol.* 53, 305 (1977).
34. F. X. Heinz and C. Kunz, *J. Gen. Virol.* 62, 271 (1982).
35. We propose an alternative nomenclature for flavivirus nonstructural proteins; our proposal is based on the yellow fever virus gene order determined by nucleic acid and protein sequence analysis (Fig. 1 and Table 1). The large nonstructural proteins (formerly NV3, NV4, and NV5) have been mapped and are numbered in order of appearance in the genome (5′ → 3′) with an upper case NS designation. We hypothesize that the remaining coding sequences in the nonstructural region encode several small flavivirus intracellular proteins (formerly NV1, NV1½, NV2, NV2½, and NVX), which are designated by a lower-case ns. Tentative identities with previously described flavivirus proteins are indicated by parentheses in Table 1. For alternative nomenclature see the text and (*37*). Minor virus-specific protein species that have been detected include two small glycoproteins (M_r ~13,000 and 17,000) (*27*, *28*), and NV4½ (apparently related to NV4; perhaps equivalent to ns2b + NS3) (*5*).
36. J. Pata, J. Schlesinger, R. Aebersold, D. Teplow, S. Kent, J. H. Strauss, C. M. Rice, unpublished data.
37. E. G. Westaway *et al.*, *Intervirology* 14, 114 (1980).
38. J. Schlesinger, personal communication.
39. G. Cleaves, personal communication.
40. R. Grantham, C. Gautier, M. Gouy, M. Jacobzone, R. Mercier, *Nucleic Acids Res.* 9, 43 (1981).
41. T. C. Hall, in *International Review of Cytology*, (Academic Press, New York, 1979), vol. 60, pp. 1–26.
42. L. S. Loesch-Fries and T. C. Hall, *Nature (London)* 298, 771 (1982); T. C. Hall, D. S. Shih, P. Kaesberg, *Biochem. J.* 129, 969 (1972).
43. P. Ahlquist, J. J. Bujarski, P. Kaesberg, T. C. Hall, *Plant Mol. Biol.* 3, 37 (1984).
44. E. G. Strauss and J. H. Strauss, *Curr. Top. Microbiol. Immunol.* 105, 1 (1983).
45. J. Haseloff, J. Haseloff, P. Goelet, D. Zimmern, P. Ahlquist, R. Dasgupta, P. Kaesberg, *Proc. Natl. Acad. Sci. U.S.A.* 81, 4358 (1984); P. Ahlquist *et al.*, *J. Virol.* 53, 536 (1985).
46. G. Kamer and P. Argos, *Nucleic Acids Res.* 12, 7269 (1984).
47. C. M. Rice and J. H. Strauss, *J. Mol. Biol.* 150, 315 (1981).
48. G. J. Russell, P. M. B. Walker, R. A. Elton, J. H. Subak-Sharpe, *ibid.* 108, 1 (1976); A. P. Bird, *Nucleic Acids Res.* 8, 1499 (1980).
49. C. M. Rice, E. G. Strauss, J. H. Strauss, in *The Togaviruses and Flaviviruses*, S. Schlesinger and M. Schlesinger, Eds. (Plenum, New York, in press).
50. Obtained from Dr. Dennis Trent, Centers for Disease Control, Fort Collins, Colorado.
51. C. M. Rice, L. Dalgarno, D. W. Trent, J. H. Strauss; details of cloning and sequencing procedures are in preparation.
52. H. Okayama and P. Berg, *Mol. Cell. Biol.* 2, 161 (1982).
53. A. M. Maxam and W. Gilbert, *Methods Enzymol.* 65, 499 (1980).
54. J. Kyte and R. F. Doolittle, *J. Mol. Biol.* 157, 105 (1982).
55. I. Tinoco *et al.*, *Nature (London)* 246, 40 (1973).

56. We thank E. G. Strauss, L. Dalgarno, and C. Chang for helpful discussions and our many colleagues for critical comments on the manuscript; L. Hood and T. Hunkapiller for the use of their computer facilities; and C. S. Hahn for help with RNA secondary structure analysis. We also thank G. Cleaves, J. J. Schlesinger, and F. X. Heinz for allowing us to quote their unpublished results. Supported in part by grants AI 20612 and AI 10793 from NIH and by grant PCM 83-16856 from NSF.

17 May 1985; accepted 7 July 1985

(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) Publication number: **0 388 232 A1**

(12) EUROPEAN PATENT APPLICATION

(21) Application number: 90302866.0

(22) Date of filing: 16.03.90

(51) Int. Cl.5: C12N 15/51, A61K 39/29, G01N 33/576, C12Q 1/70

(30) Priority: 17.03.89 US 325338
20.04.89 US 341334
18.05.89 US 355002

(43) Date of publication of application:
19.09.90 Bulletin 90/38

(84) Designated Contracting States:
AT BE CH DE DK ES FR GB GR IT LI LU NL SE

(71) Applicant: CHIRON CORPORATION
4560 Horton Street
Emeryville California 94608(US)

(72) Inventor: Houghton, Michael
53 Rosemead Court
Danville, California 94526(US)
Inventor: Choo, Qui-Lim
5700 Fern Street
El Cerrito, California 94530(US)
Inventor: Kuo, George
1370 Sixth Avenue
San Francisco, California 94122(US)

(74) Representative: Goldin, Douglas Michael et al
J.A. KEMP & CO. 14, South Square Gray's Inn
London WC1R 5EU(GB)

(54) NANBV diagnostics and vaccines.

(57) A new virus, Hepatitis C virus (HCV), which has proven to be the major etiologic agent of blood-borne NANBH, was discovered by Applicant. The initial work on this virus, which includes a partial genomic sequence of the prototype HCV isolate, is described in EPO Pub. No. 318,216, and PCT Pub. No. WO/89/04669. The present invention, which in part is based on new HCV sequences and polypeptides which are not disclosed in the above-cited publications, includes the application of these new sequences and polypeptides in immunoassays, probe diagnostics, anti-HCV antibody production, PCR technology, and recombinant DNA technology. Included within the invention also are novel immunogenic polypeptides encoded within clones containing HCV cDNA, novel methods for purifying an immuncgenic HCV polypeptide, and antisense polynucleotides derived from HCV cDNA.

EP 0 388 232 A1

conditions described supra. for the expression of the fusion polypeptide, C100-3. The resulting polypeptides are screened using the sera from individuals with NANBH, described supra. for the screening of immunogenic polypeptides encoded in HCV cDNAs expressed in E. coli.

Comparison of the Hydrophobic Profiles of HCV Polyproteins with West Nile Virus Polyprotein and with Dengue Virus NS1

The hydrophobicity profile of an HCV polyprotein segment was compared with that of a typical Flavivirus, West Nile virus. The polypeptide sequence of the West Nile virus polyprotein was deduced from the known polynucleotide sequences encoding the non-structural proteins of that virus. The HCV polyprotein sequence was deduced from the sequence of overlapping cDNA clones. The profiles were determined using an antigen program which uses a window of 7 amino acid width (the amino acid in question, and 3 residues on each side) to report the average hydrophobicity about a given amino acid residue. The parameters giving the reactive hydrophobicity for each amino acid residue are from Kyte and Doolittle (1982). Fig. 19 shows the hydrophobic profiles of the two polyproteins; the areas corresponding to the non-structural proteins of West Nile virus, ns1 through ns5, are indicated in the figure. As seen in the figure, there is a general similarity in the profiles of the HCV polyprotein and the West Nile virus polyprotein.

The sequence of the amino acids encoded in the 5′-region of HCV cDNA shown in Fig. 16 has been compared with the corresponding region of one of the strains of Dengue virus, described supra., with respect to the profile of regions of hydrophobicity and hydrophilicity (data not shown). This comparison indicated that the polypeptides from HCV and Dengue encoded in this region, which corresponds to the region encoding NS1 (or a portion thereof), have a similar hydrophobic/hydrophilic profile.

The similarity in hydrophobicity profiles, in combination with the previously identified homologies in the amino acid sequences of HCV and Dengue Flavivirus in EP 0,218,316 suggests that HCV is related to these members of the Flavivirus family.

Characterization of the Putative Polypeptides Encoded Within the HCV ORF

The sequence of the HCV cDNA sense strand, shown in Fig. 17, was deduced from the overlapping HCV cDNAs in the various clones described in EPO Pub. No. 318,216 and those described supra. It may be deduced from the sequence that the HCV genome contains primarily one long continuous ORF, which encodes a polyprotein. In the sequence, nucleotide number 1 corresponds to the first nucleotide of the initiator MET codon; minus numbers indicate that the nucleotides are that distance away in the 5′-direction (upstream), while positive numbers indicate that the nucleotides are that distance away in the 3′-direction (downstream). The composite sequence shows the "sense" strand of the HCV cDNA.

The amino acid sequence of the putative HCV polyprotein deduced from the HCV cDNA sense strand sequence is also shown in Fig. 17, where position 1 begins with the putative initiator methionine.

Possible protein domains of the encoded HCV polyprotein, as well as the approximate boundaries, are the following (the polypeptides identified within the parentheses are those which are encoded in the Flavivirus domain):

Putative Domain	Approximate Boundary (amino acid nos.)
"C" (nucleocapsid protein)	1-120
"E" (Virion envelope protein(s) and possibly matrix (M) proteins	120-400
"NS1" (complement fixation antigen?)	400-660
"NS2" (unknown function)	660-1050
"NS3" (protease?)	1050-1640
"NS4" (unknown function)	1640-2000
"NS5" (polymerase)	2000-? end

It should bo noted, however, that hydrophobicity profiles (described infra), indicate that HCV diverges from the Flavivirus model, particularly with respect to the region upstream of NS2. Moreover, the boundaries indicated are not intended to show firm demarcations between the putative polypeptides.

The Hydrophilic and Antigenic Profile of the Polypeptide

Profiles of the hydrophilicity/hydrophobicity and the antigenic index of the putative polyprotein encoded in the HCV cDNA sequence shown in Fig. 16 were determined by computer analysis. The program for hydrophilicity/hydrophobicity was as described supra. The antigenic index results from a computer program which relies on the following criteria: 1) surface probability, 2) prediction of alpha-helicity by two different methods; 3) prediction of beta-sheet regions by two different methods; 4) prediction of U-turns by two different methods; 5) hydrophilicity/hydrophobicity; and flexibility. The traces of the profiles generated by the computer analyses are shown in Fig. 20. In the hydrophilicity profile, deflection above the abscissa indicates hydrophilicity, and below the abscissa indicates hydrophobicity. The probability that a polypeptide region is antigenic is usually considered to increase when there is a deflection upward from the abscissa in the hydrophilic and/or antigenic profile. It should be noted, however, that these profiles are not necessarily indicators of the strength of the immunogenicity of a polypeptide.

Identification of Co-linear Peptides in HCV and Flaviviruses

The amino acid sequence of the putative polyprotein encoded in the HCV cDNA sense strand was compared with the known amino acid sequences of several members of Flaviviruses. The comparison shows that homology is slight, but due to the regions in which it is found, it is probably significant. The conserved colinear regions are shown in Fig. 21. The amino acid numbers listed below the sequences represent the number in the putative HCV polyprotein (See Fig. 17.)

The spacing of these conserved motifs is similar between the Flaviviruses and HCV, and implies that there is some similarity between HCV and these flaviviral agents.

The following listed materials are on deposit under the terms of the Budapest Treaty with the American Type Culture Collection (ATCC), 12301 Parklawn Dr., Rockville, Maryland 20852, and have been assigned the following Accession Numbers.

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 4 :
A61K 39/12, C12Q 1/70
C12N 15/00, C07K 7/10
A61K 37/04, 37/02, C07H 15/12

A1

(11) International Publication Number: WO 89/ 04669

(43) International Publication Date: 1 June 1989 (01.06.89)

(21) International Application Number: PCT/US88/04125

(22) International Filing Date: 18 November 1988 (18.11.88)

(31) Priority Application Numbers:
122,714
139,886
161,072
191,263
Not furnished
Not furnished

(32) Priority Dates:
18 November 1987 (18.11.87)
30 December 1987 (30.12.87)
26 February 1988 (26.02.88)
6 May 1988 (06.05.88)
26 October 1988 (26.10.88)
14 November 1988 (14.11.88)

(33) Priority Country: US

(60) Parent Application or Grant
(63) Related by Continuation
US
Filed on
Not furnished (CIP)
14 November 1988 (14.11.88)

(71) Applicant *(for all designated States except US)*: CHIRON CORPORATION [US/US]; 4560 Horton Street, Emeryville, CA 94608 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)* : HOUGHTON, Michael [GB/US]; 53 Rosemead Court, Danville, CA 94526 (US). CHOO, Qui-Kim [SG/US]; 5700 Fern Street, El Cerrito, CA 94530 (US). KUO, George [US/US]; 1370 Sixth Avenue, San Francisco, CA 94122 (US).

(74) Agents: MONROY, Gladys, H. et al.; Ciotti & Murashige, Irell & Manella, 545 Middlefield Road, Suite 200, Menlo Park, CA 94025 (US).

(81) Designated States: AT, AT (European patent), AU, BB, BE (European patent), BG, BJ (OAPI patent), BR, CF (OAPI patent), CG (OAPI patent), CH, CH (European patent), CM (OAPI patent), DE, DE (European patent), DK, FI, FR (European patent), GA (OAPI patent), GB, GB (European patent), HU, IT (European patent), JP, KP, KR, LK, LU, LU (European patent), MC, MG, ML (OAPI patent), MR (OAPI patent), MW, NL, NL (European patent), NO, RO, SD, SE, SE (European patent), SN (OAPI patent), SU, TD (OAPI patent), TG (OAPI patent), US.

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: NANBV DIAGNOSTICS AND VACCINES

(57) Abstract

A family of cDNA sequences derived from hepatitis C virus (HCV) are provided. These sequences encode antigens which react immunologically with antibodies present in individuals with non-A non-B hepatitis (NANBH), but which generally are absent from individuals infected with hepatitis A virus (HAV) or hepatitis B virus (HBV), and also are absent from control individuals. A comparison of these cDNA sequences with the sequences in Genebank, and with the sequences of hepatitis delta virus (HDV) and HBV shows a lack of substantial homology. A comparison of the sequences of amino acids encoded in the cDNA with the sequences of Flaviviruses indicates that HCV is a Flavivirus or Flavi-like virus. The HCV cDNA sequences are useful for the design of polynucleotide probes, and for the synthesis of polypeptides which may be used in immunoassays. Both the polynucleotide probes and the polypeptides may be useful for the diagnosis of HCV-induced NANBH, and for screening blood bank specimens and donors for HCV infection. In addition, these cDNA sequences may be useful for the synthesis of immunogenic polypeptides which may be used in vaccines for the treatment, prophylactic and/or therapeutic, of HCV infection. Polypeptides encoded within the cDNA sequences may also be used to raise antibodies against HCV antigenes, and for the purification of antibodies directed against HCV antigenes. These antibodies may be useful in immunoassays for detecting HCV antigenes associated with NANBH in individuals, and in blood bank donations. Moreover, these antibodies may be used for treatment of NANBH in individuals. The reagents provided in the invention also enable the isolation of NANBH agent(s), and the propagation of these agent(s) in tissue culture systems. Moreover, they provide reagents which are useful for screening for antiviral agents for HCV, particularly in tissue culture or animal model systems.

carried out utilizing the PCR amplification procedure, as described in Section IV.C.3, except that the hybridization probe was a kinased oligonucleotide derived from the clone 81 cDNA sequence. The results showed that the amplified sequences hybridized with the clone 81 derived HCV cDNA probe.

IV.H.3. Homology Between the Non-Structural Protein of Dengue Flavivirus (MNWWVD1) and the HCV Polypeptides Encoded by the Combined ORF of Clones 14i Through 39c

The combined HCV cDNAs of clones 14i through 39c contain one continuous ORF, as shown in Fig. 26. The polypeptide encoded therein was analyzed for sequence homology with the region of the non-structural polypeptide(s) in Dengue flavivirus (MNWVD1). The analysis used the Dayhoff protein data base, and was performed on a computer. The results are shown in Fig. 42, where the symbol (:) indicates an exact homology, and the symbol (.) indicates a conservative replacement in the sequence; the dashes indicate spaces inserted into the sequence to achieve the greatest homologies. As seen from the figure, there is significant homology between the sequence encoded in the HCV cDNA, and the non-structural polypeptide(s) of Dengue virus. In addition to the homology shown in Fig. 42, analysis of the polypeptide segment encoded in a region towards the 3'-end of the cDNA also contained sequences which are homologous to sequences in the Dengue polymerase. Of consequence is the finding that the canonical Gly-Asp-Asp (GDD) sequence thought to be essential for RNA-dependent RNA polymerases is contained in the polypeptide encoded in HCV cDNA, in a location which is consistent with that in Dengue 2 virus. (Data not shown.)

(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) Publication number: **0 318 216**
A1

(12) # EUROPEAN PATENT APPLICATION

(21) Application number: 88310922.5

(22) Date of filing: 18.11.88

(51) Int. Cl.4: C12N 15/00 , C12N 7/00 , A61K 39/29 , G01N 33/576 , C12Q 1/70 , C12Q 1/68 , C07K 15/00 , A61K 39/395 , C12P 21/00

(30) Priority: 18.11.87 US 122714
30.12.87 US 139886
26.02.88 US 161072
06.05.88 US 191263
26.10.88 US 263584
14.11.88 US 271450

(43) Date of publication of application:
31.05.89 Bulletin 89/22

(84) Designated Contracting States:
AT BE CH DE ES FR GB GR IT LI LU NL SE

(71) Applicant: CHIRON CORPORATION
4560 Horton Street
Emeryville California 94608(US)

(72) Inventor: Houghton, Michael
1307 Greenbrook Drive
Danville California 94526(US)
Inventor: Choo, qui-Lim
5700 Fern Street
El Cerrito California 94530(US)
Inventor: Kuo, George
1370 Sixth Avenue
San Francisco California 94122(US)

(74) Representative: Goldin, Douglas Michael et al
J.A. KEMP & CO. 14, South Square Gray's Inn
London WC1R 5EU(GB)

(54) **Nanbv diagnostics and vaccines.**

(57) A family of cDNA sequences derived from hepatitis C virus (HCV) are provided. These sequences encode antigens which react immunologically with antibodies present in individuals with non-A non-B hepatitis (NANBH), but which generally are absent from individuals infected with hepatitis A virus (HAV) or hepatitis B virus (HBV), and also are absent from control individuals. A comparison of these cDNA sequences with the sequences in Genebank, and with the sequences of hepatitis delta virus (HDV) and HBV shows a lack of substantial homology. A comparison of the sequences of amino acids encoded in the cDNA with the sequences of Flaviviruses indicates that HCV is a Flavivirus or Flavi-like virus.

The HCV cDNA sequences are useful for the design of polynucleotide probes, and for the synthesis of polypeptides which may be used in immunoassays. Both the polynucleotide probes and the polypeptides may be useful for the diagnosis of HCV-induced NANBH, and for screening blood bank specimens and donors for HCV infection. In addition, these cDNA sequences may be useful for the synthesis of immunogenic polypeptides which may be used in vaccines for the treatment, prophylactic and/or therapeutic, of HCV infection. Polypeptides encoded within the cDNA sequences may also be used to raise antibodies against HCV antigens, and for the purification of antibodies directed against HCV antigens. These antibodies may be useful in immunoassays for detecting HCV antigens associated with NANBH in individuals, and in blood bank donations. Moreover, these antibodies may be used for treatment of NANBH in individuals.

The reagents provided in the invention also enable the isolation of NANBH agent(s), and the propagation of

EP 0 318 216 A1

these agent(s) in tissue culture systems. Moreover, they provide reagents which are useful for screening for antiviral agents for HCV, particularly in tissue culture or animal model systems.

from clones 81, 40b, and 25c. This mixture was used to increase the sensitivity of the hybridization assay. The samples in panel I were hybridized with the plus strand probe mixture. The samples in panel II were probed by hybridization with the minus strand probe mixture. The composition of the samples in the panels of the immunoblot are presented in table 4.

Table 4

lane	A	B
1	HCV genome	*
2	---	*
3	*	cDNA 81
4	---	cDNA 81

* is an undescribed sample.

As seen from the results in Fig. 41, only the minus strand DNA probe hybridizes with the isolated HCV genome. This result, in combination with the result showing that the genome is sensitive to RNase and not DNase (See Section IV.C.2.), suggests that the genome of NANBV is positive stranded RNA.

These data, and data from other laboratories concerning the physicochemical properties of a putative NANBV(s), are consistent with the possibility that HCV is a member of the Flaviviridae. However, the possibility that HCV represents a new class of viral agent has not been eliminated.

IV.H.2. Detection of Sequences in Captured Particles Which When Amplified by PCR Hybridize to HCV cDNA Derived from Clone 81

The RNA in captured particles was obtained as described in Section IV.H.1. The analysis for sequences which hybridize to the HCV cDNA derived from clone 81 was carried out utilizing the PCR amplification procedure, as described in Section IV.C.3, except that the hybridization probe was a kinased oligonucleotide derived from the clone 81 cDNA sequence. The results showed that the amplified sequences hybridized with the clone 81 derived HCV cDNA probe.

IV.H.3. Homology Between the Non-Structural Protein of Dengue Flavivirus (MNWWVD1) and the HCV Polypeptides Encoded by the Combined ORF of Clones 14i Through 39c

The combined HCV cDNAs of clones 14i through 39c contain one continuous ORF, as shown in Fig. 26. The polypeptide encoded therein was analyzed for sequence homology with the region of the non-structural polypeptide(s) in Dengue flavivirus (MNWVD1). The analysis used the Dayhoff protein data base, and was performed on a computer. The results are shown in Fig. 42, where the symbol (:) indicates an exact homology, and the symbol (.) indicates a conservative replacement in the sequence; the dashes indicate spaces inserted into the sequence to achieve the greatest homologies. As seen from the figure, there is significant homology between the sequence encoded in the HCV cDNA, and the non-structural polypeptide-(s) of Dengue virus. In addition to the homology shown in Fig. 42, analysis of the polypeptide segment encoded in a region towards the 3'-end of the cDNA also contained sequences which are homologous to sequences in the Dengue polymerase. Of consequence is the finding that the canonical Gly-Asp-Asp (GDD) sequence thought to be essential for RNA-dependent RNA polymerases is contained in the polypeptide encoded in HCV cDNA, in a location which is consistent with that in Dengue 2 virus. (Data not shown.)

IV.H.4. HCV-DNA is Not Detectable in NANBH Infected Tissue

Two types of studies provide results suggesting that HCV-DNA is not detectable in tissue from an individual with NANBH. These results, in conjunction with those described in IV.C. and IV.H.1. and IV.H.2. provide evidence that HCV is not a DNA containing virus, and that its replication does not involve cDNA.

IV.H.4.a. Southern Blotting Procedure

FIG. 41-1

Homology between the HCV polypeptide encoded by combined ORF of clones 14i through 39c) and the non-structural protein of the Dengue flavivirus(MNWVD1).

```
                      10        20        30        40        50
HCV         EYVVLLFLLLADARVCSCLWMMLLISQAEAALENLVILNAASLAGTHGLVSFLVFFCFA

MNWVD1     AVSFVTLITGNMSFRDLGRVMVMVGATMTDDIGMGVTYLALLAAFKVRPTFAAGLLLRKL
                130       140       150       160       170       180

          60        70        80        90       100       110
HCV        WYLKGKWVPGAVYTFYGMWPLLLLLLALPQRAYALDTEVAASCGGVVLVGLMALTLSPYY

MNWVD1     TSKELMMTTIGIVLLSQSTIPETILELTDALALGMMVLKMVRKMEKYQLAVTIMAILCVP
                190       200       210       220       230       240

         120       130       140       150       160       170
HCV        KRYISWCLWWLQYFLTRVEAQLHVWIPPLNVRGGRDAVILLMCAVHPTLVFDITKLLLAV
                  :  .  .  .  :  .  ...... . .... .. :  . ...:: .: .:.:  .
MNWVD1     NAVILQNAWKVSCTILAVVSVSPLFLTSSQQKADWIPLALTIKGLNPTAIF-LTTLSRTN
                250       260       270       280       290

         180       190       200       210       220       230
HCV        FGPLWILQASLLKVPYF-VRVQGLLRF-CALARKMIGGHYVQMVIIKLGALTGTYVYNHL
             . : :..... :    . . .::.   ...   ...:  .  .  . :.. ...   ..
MNWVD1     KKRSWPLNEAIMAVGMVSILASSLLKNDIPMTGPLVAGGLLTVCYV-LTGRSADLELERA
         300       310       320       330       340       350

          240       250       260       270       280       290
HCV        TPLRDWAHNGLRDLAVAVEPVVFSQMETKLITWGADTAACGDIINGLPVSARRGREILLG
           .  .. :....  .  .  .. ....:. .... . ......  .:.    .  .  .:    :.
MNWVD1     ADVK-WEDQAEISGSSPILSITISE-DGSMSIKNEEEEQTLTILIRTGLLVISG---LFP
          360       370       380       390       400       410

          300       310       320       330       340       350
HCV        PADGMVSKGWRLLAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATC
            .  ... .: :  .  .. :   ... .  .. .. ....   .  .  .  .  .:.  ...
MNWVD1     VSIPITAAAWYLWEVKKQRAGVLWDVPSPPPVGKAELEDGAYRIKQKGILGYSQIGAGVY
              420       430       440       450       460       470

          360       370       380       390                400      410
HCV        INGVCWTVYHGAGTRTIASPKGPVIQMYTNVDQDLV-----GWPAPQGSRSLTPCTCGSSD
            .:.  :..: .  .  ..       .. .  ...:..:::    ::      .  ..  .     . .
MNWVD1     KEGTFHTMWHVTRGAVLMHKGKRIEPSWADVKKDLVSCGGGWKLEGEWKEGEEVQVLALE
              480       490       500       510       520       530

           420       430       440       450       460       470
HCV        LYLVTRHADVIPVRRRGDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGV
             .: ... :.  ...   .. ...  ....  :.::.:..   :..::..  ..: ::.
MNWVD1     PGKNPRAVQTKPGLFKTN--AGTIGAVSLDFSPGTSGSPIIDKKGKVVGLYGNGVVTRSG
              540       550       560       570       580       590

           480       490       500       510       520          530
HCV        AKAVDFIPVENLETTMRSPVFTDNSSPPVVPQSFQVAHLHAPTGSGKS--TKVPAAYAAQ
           :  .  ..  ..:.  ..   .: ..:.  .    ... .  .::.  .:..:.   .  .  .  .
MNWVD1     AYVSAIAQTEK--SIEDNPEIEDDIFRK---RKLTIMDLHPGAGKTKRYLPAIVRGAIKR
               600       610        620       630       640

               540       550       560       570       580
```

```
HCV       GYKVLVLNPS--VAATLGFGAYMSKAHGIDPNIRTGVRTITTGSPITYSTYGKFLADGGC
          : ..:.:.:.  :::... .     . .:.:::. . .. . . ..:... : .
MNWVD1    GLRTLILAPTRVVAAEMEEALRGLPIRYQTPAIRAEHTGREIVDLMCHATFTMRLL-SPV
           650       660       670       680       690       700

          590       600       610       620       630       640
HCV       SGGAYDIIICDECHSTDATSILGIGTVLDQAETAGARLVVLATATPPGSVTVPHPNIEEV
          . .X..:: :: : ::..:: . : . ...: :.   .. :::::::: . : : ^ v.
MNWVD1    RVPNYNLIIMDEAHFTDPASIAARGYISTRVE-MGEAAGIFMTATPPGSRD-PFPQSNAP
            710       720       730       740       750       760

          650       660       670       680       690       700
HCV       ALSTTGEIPFYGKAIPLEVIKGGRHLIFCHSKKKCDELAAKLVALGINAVAYYRGLDVSV
           .... :::
MNWVD1    IMDEEREIPERSWSSGHEWVTDFKGKTVWFVPSIKAGNDTAACLRKNGKKVTQLSRKTFD
              770       780       790       800       810       820

          710       720       730       740       750       760
HCV       IPTSGDVVVVATDALMTGYTGDFDSVIDCNTCVTQTVDFSLDPTFTIETITLPQDAVSRT

MNWVD1    SEYVKTRTNDWNFVVTTDISEMGANFKAERVIDPRRCMKPVILTDGEERVILAGPMPVTH
              830       840       850       860       870       880

          770       780       790       800       810       820
HCV       QRRGRTGRGKPGIYRFVAPGERPSGMFDSSVLCECYDAGCAWYELTPAETTVRLRAYMNT

MNWVD1    SS
```

FIG. 41-2

ASSIGNMENT

This Assignment is made by **Tatsuo Miyamura**, residing at 4-21-22-113, Hamadayama, Suginami-ku, Tokyo 168, Japan, and **Izumu Saito**, residing at 2-37-15-412, Yoyogi, Shibuya-ku, Tokyo 151, Japan, who are hereinafter collectively referred to as "**Assignors.**"

The Assignors have made certain new and useful inventions set forth in a United States patent application entitled "**New HCV Isolate**" filed in the name of the Assignors on 15 September 1989 under attorney docket number 2300-0089, and given Serial No. 408,045 by the United States Patent and Trademark Office, hereinafter referred to as the "**Inventions**" or the "**Patent Application**", respectively.

The **Director General of the National Institute of Health of Japan**, on behalf of the National Institute of Health of Japan, having a place of business located at 2-10-35 Kamiosaki, Shinagawa-ku, Tokyo 141, Japan, and **Chiron Corporation**, a Delaware corporation having a place of business located at 4560 Horton Street, Emeryville, California 94608, United States of America, hereinafter collectively referred to as "**Assignees**", desire to collectively obtain all rights to the Inventions, the Patent Application, and any letters patent, United States or foreign, obtained therefor or thereon.

In consideration of the sum of one dollar ($1.00) and other good and valuable consideration actually received, the Assignors agree to assign, and hereby do assign, transfer and set over to the Assignees, in equal and undivided shares, the entire right, title and interest in the Inventions, the Patent Application, any and all letters patent in the United States of America and all foreign countries which may be granted therefor or thereon, any and all continuations, divisions and continuation-in-parts of the Patent Application, any and all reissues and extensions of such letters patent, and all rights under the International Convention for the Protection of Industrial Property (also known as the "Paris Convention") arising from the Inventions and the Patent Application.

For the same consideration recited above, the Assignors also:

1) agree to execute all papers necessary in connection with the Patent Application and any continuing or divisional or reissue applications thereof and also to execute separate assignments in connection with such applications as the Assignees may deem necessary or expedient or essential to its full protection and title in and to the invention hereby transferred;

2) agree to execute all papers necessary in connection with any interference which may be declared concerning the Patent Application or

continuation or division or re-issue thereof and to cooperate with the Assignees in every way possible in obtaining evidence and going forward with such interference;
3) agree to perform all affirmative acts which may be necessary to obtain a grant of a valid United States or foreign patent to the Assignees;
4) agree to communicate to the Assignees or representatives thereof any facts known to them, testify in any legal proceedings regarding the Invention;
5) authorize and request the Commissioner of Patents to issue any and all Letters Patents of the United States resulting from said application or any division or divisions or continuing applications thereof to the said Assignees, as the Assignees of the entire interest, and hereby covenants that they have full right to convey the entire interest herein assigned, and that they have not executed and will not execute, any agreement in conflict herewith; and
6) grant the firm of Irell & Manella the power to insert on this assignment any further identification which may be necessary or desirable in order to comply with the rules of the United States Patent Office for recordation of this document.

This Assignment shall be binding upon our heirs, executors, administrators, and/or assigns, and shall inure to this benefit of the heirs, executors, administrators, successors and/or assigns of the Assignees.

In witness whereof, executed by the Assignors on the date(s) indicated below.

Tatsuo Miyamura — Date: September 22, 1989
Tatsuo Miyamura

Izumu Saito — Date: Sept. 22, 1989
Izumu Saito

The above signatures were made in my presence, or acknowledged to me, by Tatsuo Miyamura and Izumu Saito, who are both known to me:

Witness H. Shimojo — Date: September 22, 1989
Name Hiroto Shimojo
(Type or Print)

(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) Publication number: **0 318 216 B1**

(12) **EUROPEAN PATENT SPECIFICATION**

(45) Date of publication of patent specification: **15.12.93**

(51) Int. Cl.5: **C12N 15/00**, C12N 7/00, A61K 39/29, G01N 33/576, C12Q 1/70, C12Q 1/68, C07K 15/00, A61K 39/395, C12P 21/00

(21) Application number: **88310922.5**

(22) Date of filing: **18.11.88**

The file contains technical information submitted after the application was filed and not included in this specification

(54) NANBV diagnostics and vaccines.

(30) Priority: **18.11.87 US 122714**
30.12.87 US 139886
26.02.88 US 161072
06.05.88 US 191263
26.10.88 US 263584
14.11.88 US 271450

(43) Date of publication of application:
31.05.89 Bulletin 89/22

(45) Publication of the grant of the patent:
15.12.93 Bulletin 93/50

(84) Designated Contracting States:
AT BE CH DE ES FR GR IT LI LU NL SE

(56) References cited:
EP-A- 0 194 207
EP-A- 0 263 761
EP-A- 0 279 460
WO-A-82/00205
WO-A-82/02774
WO-A-88/03410

(73) Proprietor: **CHIRON CORPORATION**
4560 Horton Street
Emeryville, California 94608(US)

(72) Inventor: **Houghton, Michael**
1307 Greenbrook Drive
Danville California 94526(US)
Inventor: **Choo, Qui-Lim**
5700 Fern Street
El Cerrito California 94530(US)
Inventor: **Kuo, George**
1370 Sixth Avenue
San Francisco California 94122(US)

(74) Representative: **Goldin, Douglas Michael et al**
J.A. KEMP & CO.
14, South Square
Gray's Inn
London WC1R 5EU (GB)

EP 0 318 216 B1

sequence of the polypeptide encoded in the extended ORF in the derived sequence.

Fig. 27 shows the sequence of the HCV cDNA in clone 12f, the segment which overlaps clone 14i, and the amino acids encoded therein.

Fig. 28 shows the sequence of the HCV cDNA in clone 35f, the segment which overlaps clone 39c, and the amino acids encoded therein.

Fig. 29 shows the sequence of the HCV cDNA in clone 19g, the segment which overlaps clone 35f, and the amino acids encoded therein.

Fig. 30 shows the sequence of clone 26g, the segment which overlaps clone 19g, and the amino acids encoded therein.

Fig. 31 shows the sequence of clone 15e, the segment which overlaps clone 26g, and the amino acids encoded therein.

Fig. 32 shows the sequence in a composite cDNA, which was derived by aligning clones 12f through 15e in the 5' to 3' direction; it also shows the amino acids encoded in the continuous ORF.

Fig. 33 shows a photograph of Western blots of a fusion protein, SOD-NANB$_{5-1-1}$, with chimpanzee serum from chimpanzees infected with BB-NANB, HAV, and HBV.

Fig. 34 shows a photograph of Western blots of a fusion protein, SOD-NANB$_{5-1-1}$, with serum from humans infected with NANBV, HAV, HBV, and from control humans.

Fig. 35 is a map showing the significant features of the vector pAB24.

Fig. 36 shows the putative amino acid sequence of the carboxy-terminus of the fusion polypeptide C100-3 and the nucleotide sequence encoding it.

Fig. 37A is a photograph of a coomassie blue stained polyacrylamide gel which identifies C100-3 expressed in yeast.

Fig. 37B shows a Western blot of C100-3 with serum from a NANBV infected human.

Fig. 38 shows an autoradiograph of a Northern blot of RNA isolated from the liver of a BB-NANBV infected chimpanzee, probed with BB-NANBV cDNA of clone 81.

Fig. 39 shows an autoradiograph of NANBV nucleic acid treated with RNase A or DNase I, and probed with BB-NANBV cDNA of clone 81.

Fig. 40 shows an autoradiograph of nucleic acids extracted from NANBV particles captured from infected plasma with anti-NANB$_{5-1-1}$, and probed with ^{32}P-labeled NANBV cDNA from clone 81.

Fig. 41a and b shows autoradiographs of filters containing isolated NANBV nucleic acids, probed with ^{32}P-labeled plus and minus strand DNA probes derived from NANBV cDNA in clone 81.

Fig. 41-1 shows the homologies between a polypeptide encoded in HCV cDNA and an NS protein from Dengue flavivirus.

Fig. 43 shows a histogram of the distribution of HCV infection in random samples, as determined by an ELISA screening.

Fig. 44 shows a histogram of the distribution of HCV infection in random samples using two configurations of immunoglobulin-enzyme conjugate in an ELISA assay.

Fig. 45 shows the sequences in a primer mix, derived from a conserved sequence in NS1 of flaviviruses.

Fig. 46 shows the HCV cDNA sequence in clone k9-1, the segment which overlaps the cDNA in Fig. 27, and the amino acids encoded therein.

Fig. 47 shows the sequence in a composite CDNA which was derived by aligning clones k9-1 through 15e in the 5' to 3' direction; it also shows the amino acids encoded in the continuous ORF.

I. Definitions

The term "hepatitis C virus" has been reserved by workers in the field for an heretofore unknown etiologic agent of NANBH. Accordingly, as used herein, "hepatitis C virus" (HCV) refers to an agent causitive of NANBH, which agent is a virus characterised by:(i) a positive stranded RNA genome; (ii) said genome comprising an open reading frame (ORF) encoding a polyprotein; and (iii) the portion of said polyprotein corresponding to Figure 14 having at least 40% homology to the amino acid sequence in Figure 14. This agent was formerly referred to as NANBV and/or BB-NANBV. The terms HCV, NANBV, and BB-NANBV are used interchangeably herein, but all refer to the virus as defined above. As an extension of this terminology, the disease caused by HCV, formerly Called NANB hepatitis (NANBH), is called hepatitis C. The terms NANBH and hepatitis C may be used interchangeably herein.

The term "HCV", as used herein, denotes a viral species which causes NANBH, and attenuated strains or defective interfering particles derived therefrom. As shown infra., the HCV genome is comprised of RNA. It is known that RNA containing viruses have relatively high rates of spontaneous mutation, i.e., reportedly

FIG. 32-1 COMBINED ORF OF DNAs 12f through 15e

```
      IlePheLysIleArgMetTyrValGlyGlyValGluHisArgLeuGluAlaAlaCysAsn
  1   CCATATTTAAAATCAGGATGTACGTGGGAGGGGTCGAACACAGGCTGGAAGCTGCCTGCA
      GGTATAAATTTTAGTCCTACATGCACCCTCCCCAGCTTGTGTCCGACCTTCGACGGACGT

      TrpThrArgGlyGluArgCysAspLeuGluAspArgAspArgSerGluLeuSerProLeu
 61   ACTGGACGCGGGGCGAACGTTGCGATCTGGAAGACAGGGACAGGTCCGAGCTCAGCCCGT
      TGACCTGCGCCCCGCTTGCAACGCTAGACCTTCTGTCCCTGTCCAGGCTCGAGTCGGGCA

      LeuLeuThrThrThrGlnTrpGlnValLeuProCysSerPheThrThrLeuProAlaLeu
121   TACTGCTGACCACTACACAGTGGCAGGTCCTCCCGTGTTCCTTCACAACCCTACCAGCCT
      ATGACGACTGGTGATGTGTCACCGTCCAGGAGGGCACAAGGAAGTGTTGGGATGGTCGGA

      SerThrGlyLeuIleHisLeuHisGlnAsnIleValAspValGlnTyrLeuTyrGlyVal
181   TGTCCACCGGCCTCATCCACCTCCACCAGAACATTGTGGACGTGCAGTACTTGTACGGGG
      ACAGGTGGCCGGAGTAGGTGGAGGTGGTCTTGTAACACCTGCACGTCATGAACATGCCCC

      GlySerSerIleAlaSerTrpAlaIleLysTrpGluTyrValValLeuLeuPheLeuLeu
241   TGGGGTCAAGCATCGCGTCCTGGGCCATTAAGTGGGAGTACGTCGTTCTCCTGTTCCTTC
      ACCCCAGTTCGTAGCGCAGGACCCGGTAATTCACCCTCATGCAGCAAGAGGACAAGGAAG

      LeuAlaAspAlaArgValCysSerCysLeuTrpMetMetLeuLeuIleSerGlnAlaGlu
301   TGCTTGCAGACGCGCGCGTCTGCTCCTGCTTGTGGATGATGCTACTCATATCCCAAGCGG
      ACGAACGTCTGCGCGCGCAGACGAGGACGAACACCTACTACGATGAGTATAGGGTTCGCC

      AlaAlaLeuGluAsnLeuValIleLeuAsnAlaAlaSerLeuAlaGlyThrHisGlyLeu
361   AGGCGGCTTTGGAGAACCTCGTAATACTTAATGCAGCATCCCTGGCCGGGACGCACGGTC
      TCCGCCGAAACCTCTTGGAGCATTATGAATTACGTCGTAGGGACCGGCCCTGCGTGCCAG

      ValSerPheLeuValPhePheCysPheAlaTrpTyrLeuLysGlyLysTrpValProGly
421   TTGTATCCTTCCTCGTGTTCTTCTGCTTTGCATGGTATTTGAAGGGTAAGTGGGTGCCCG
      AACATAGGAAGGAGCACAAGAAGACGAAACGTACCATAAACTTCCCATTCACCCACGGGC

      AlaValTyrThrPheTyrGlyMetTrpProLeuLeuLeuLeuLeuAlaLeuProGln
481   GAGCGGTCTACACCTTCTACGGGATGTGGCCTCTCCTCCTGCTCCTGTTGGCGTTGCCCC
      CTCGCCAGATGTGGAAGATGCCCTACACCGGAGAGGAGGACGAGGACAACCGCAACGGGG

      ArgAlaTyrAlaLeuAspThrGluValAlaAlaSerCysGlyGlyValValLeuValGly
541   AGCGGGCGTACGCGCTGGACACGGAGGTGGCCGCGTCGTGTGGCGGTGTTGTTCTCGTCG
      TCGCCCGCATGCGCGACCTGTGCCTCCACCGGCGCAGCACACCGCCACAACAAGAGCAGC

      LeuMetAlaLeuThrLeuSerProTyrTyrLysArgTyrIleSerTrpCysLeuTrpTrp
601   GGTTGATGGCGCTGACTCTGTCACCATATTACAAGCGCTATATCAGCTGGTGCTTGTGGT
      CCAACTACCGCGACTGAGACAGTGGTATAATGTTCGCGATATAGTCGACCACGAACACCA

      LeuGlnTyrPheLeuThrArgValGluAlaGlnLeuHisValTrpIleProProLeuAsn
661   GGCTTCAGTATTTTCTGACCAGAGTGGAAGCGCAACTGCACGTGTGGATTCCCCCCCTCA
      CCGAAGTCATAAAAGACTGGTCTCACCTTCGCGTTGACGTGCACACCTAAGGGGGGGAGT

      ValArgGlyGlyArgAspAlaValIleLeuLeuMetCysAlaValHisProThrLeuVal
721   ACGTCCGAGGGGGGCGCGACGCCGTCATCTTACTCATGTGTGCTGTACACCCGACTCTGG
      TGCAGGCTCCCCCCGCGCTGCGGCAGTAGAATGAGTACACACGACATGTGGGCTGAGACC

      PheAspIleThrLysLeuLeuLeuAlaValPheGlyProLeuTrpIleLeuGlnAlaSer
781   TATTTGACATCACCAAATTGCTGCTGGCCGTCTTCGGACCCCTTTGGATTCTTCAAGCCA
      ATAAACTGTAGTGGTTTAACGACGACCGGCAGAAGCCTGGGGAAACCTAAGAAGTTCGGT

      LeuLeuLysValProTyrPheValArgValGlnGlyLeuLeuArgPheCysAlaLeuAla
841   GTTTGCTTAAAGTACCCTACTTTGTGCGCGTCCAAGGCCTTCTCCGGTTCTGCGCGTTAG
      CAAACGAATTTCATGGGATGAAACACGCGCAGGTTCCGGAAGAGGCCAAGACGCGCAATC

      ArgLysMetIleGlyGlyHisTyrValGlnMetValIleIleLysLeuGlyAlaLeuThr
901   CGCGGAAGATGATCGGAGGCCATTACGTGCAAATGGTCATCATTAAGTTAGGGGCGCTTA
      GCGCCTTCTACTAGCCTCCGGTAATGCACGTTTACCAGTAGTAATTCAATCCCCGCGAAT
```

```
          GlyThrTyrValTyrAsnHisLeuThrProLeuArgAspTrpAlaHisAsnGlyLeuArg
961   CTGGCACCTATGTTTATAACCATCTCACTCCTCTTCGGGACTGGGCGCACAACGGCTTGC
      GACCGTGGATACAAATATTGGTAGAGTGAGGAGAAGCCCTGACCCGCGTGTTGCCGAACG

          AspLeuAlaValAlaValGluProValValPheSerGlnMetGluThrLysLeuIleThr
1021  GAGATCTGGCCGTGGCTGTAGAGCCAGTCGTCTTCTCCCAAATGGAGACCAAGCTCATCA
      CTCTAGACCGGCACCGACATCTCGGTCAGCAGAAGAGGGTTTACCTCTGGTTCGAGTAGT

          TrpGlyAlaAspThrAlaAlaCysGlyAspIleIleAsnGlyLeuProValSerAlaArg
1081  CGTGGGGGGCAGATACCGCCGCGTGCGGTGACATCATCAACGGCTTGCCTGTTTCCGCCC
      GCACCCCCCGTCTATGGCGGCGCACGCCACTGTAGTAGTTGCCGAACGGACAAAGGCGGG

          ArgGlyArgGluIleLeuLeuGlyProAlaAspGlyMetValSerLysGlyTrpArgLeu
1141  GCAGGGGCCGGGAGATACTGCTCGGGCCAGCCGATGGAATGGTCTCCAAGGGGTGGAGGT
      CGTCCCCGGCCCTCTATGACGAGCCCGGTCGGCTACCTTACCAGAGGTTCCCCACCTCCA

          LeuAlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThr
1201  TGCTGGCGCCCATCACGGCGTACGCCCAGCAGACAAGGGGCCTCCTAGGGTGCATAATCA
      ACGACCGCGGGTAGTGCCGCATGCGGGTCGTCTGTTCCCCGGAGGATCCCACGTATTAGT

          SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAla
1261  CCAGCCTAACTGGCCGGGACAAAAACCAAGTGGAGGGTGAGGTCCAGATTGTGTCAACTG
      GGTCGGATTGACCGGCCCTGTTTTTGGTTCACCTCCCACTCCAGGTCTAACACAGTTGAC

          AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAla
1321  CTGCCCAAACCTTCCTGGCAACGTGCATCAATGGGGTGTGCTGGACTGTCTACCACGGGG
      GACGGGTTTGGAAGGACCGTTGCACGTAGTTACCCCACACGACCTGACAGATGGTGCCCC

          GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAsp
1381  CCGGAACGAGGACCATCGCGTCACCCAAGGGTCCTGTCATCCAGATGTATACCAATGTAG
      GGCCTTGCTCCTGGTAGCGCAGTGGGTTCCCAGGACAGTAGGTCTACATATGGTTACATC

          GlnAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThrCys
1441  ACCAAGACCTTGTGGGCTGGCCCGCTCCGCAAGGTAGCCGCTCATTGACACCCTGCACTT
      TGGTTCTGGAACACCCGACCGGGCGAGGCGTTCCATCGGCGAGTAACTGTGGGACGTGAA

          GlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArgArg
1501  GCGGCTCCTCGGACCTTTACCTGGTCACGAGGCACGCCGATGTCATTCCCGTGCGCCGGC
      CGCCGAGGAGCCTGGAAATGGACCAGTGCTCCGTGCGGCTACAGTAAGGGCACGCGGCCG

          GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSer
1561  GGGGTGATAGCAGGGGCAGCCTGCTGTCGCCCCGGCCCATTTCCTACTTGAAAGGCTCCT
      CCCCACTATCGTCCCCGTCGGACGACAGCGGGGCCGGGTAAAGGATGAACTTTCCGAGGA

          GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCys
1621  CGGGGGGTCCGCTGTTGTGCCCCGCGGGGCACGCCGTGGGCATATTTAGGGCCGCGGTGT
      GCCCCCCAGGCGACAACACGGGGCGCCCCGTGCGGCACCCGTATAAATCCCGGCGCCACA

          ThrArgGlyValAlaLysAlaValAspPheIleProValGluAsnLeuGluThrThrMet
1681  GCACCCGTGGAGTGGCTAAGGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGACAACCA
      CGTGGGCACCTCACCGATTCCGCCACCTGAAATAGGGACACCTCTTGGATCTCTGTTGGT

          ArgSerProValPheThrAspAsnSerSerProProValValProGlnSerPheGlnVal
1741  TGAGGTCCCCGGTGTTCACGGATAACTCCTCTCCACCAGTAGTGCCCCAGAGCTTCCAGG
      ACTCCAGGGGCCACAAGTGCCTATTGAGGAGAGGTGGTCATCACGGGGTCTCGAAGGTCC

          AlaHisLeuHisAlaProThrGlySerGlyLysSerThrLysValProAlaAlaTyrAla
1801  TGGCTCACCTCCATGCTCCCACAGGCAGCGGCAAAAGCACCAAGGTCCCGGCTGCATATG
      ACCGAGTGGAGGTACGAGGGTGTCCGTCGCCGTTTTCGTGGTTCCAGGGCCGACGTATAC

          AlaGlnGlyTyrLysValLeuValLeuAsnProSerValAlaAlaThrLeuGlyPheGly
1861  CAGCTCAGGGCTATAAGGTGCTAGTACTCAACCCCTCTGTTGCTGCAACACTGGGCTTTG
      GTCGAGTCCCGATATTCCACGATCATGAGTTGGGGAGACAACGACGTTGTGACCCGAAAC

          AlaTyrMetSerLysAlaHisGlyIleAspProAsnIleArgThrGlyValArgThrIle
```

FIG. 32-2

```
1921  GTGCTTACATGTCCAAGGCTCATGGGATCGATCCTAACATCAGGACCGGGGTGAGAACAA
      CACGAATGTACAGGTTCCGAGTACCCTAGCTAGGATTGTAGTCCTGGCCCCACTCTTGTT

        ThrThrGlySerProIleThrTyrSerThrTyrGlyLysPheLeuAlaAspGlyGlyCys
1981  TTACCACTGGCAGCCCCATCACGTACTCCACCTACGGCAAGTTCCTTGCCGACGGCGGGT
      AATGGTGACCGTCGGGGTAGTGCATGAGGTGGATGCCGTTCAAGGAACGGCTGCCGCCCA

        SerGlyGlyAlaTyrAspIleIleIleCysAspGluCysHisSerThrAspAlaThrSer
2041  GCTCGGGGGGCGCTTATGACATAATAATTTGTGACGAGTGCCACTCCACGGATGCCACAT
      CGAGCCCCCCGCGAATACTGTATTATTAAACACTGCTCACGGTGAGGTGCCTACGGTGTA

        IleLeuGlyIleGlyThrValLeuAspGlnAlaGluThrAlaGlyAlaArgLeuValVal
2101  CCATCTTGGGCATCGGCACTGTCCTTGACCAAGCAGAGACTGCGGGGGCGAGACTGGTTG
      GGTAGAACCCGTAGCCGTGACAGGAACTGGTTCGTCTCTGACGCCCCCGCTCTGACCAAC

        LeuAlaThrAlaThrProProGlySerValThrValProHisProAsnIleGluGluVal
2161  TGCTCGCCACCGCCACCCCTCCGGGCTCCGTCACTGTGCCCCATCCCAACATCGAGGAGG
      ACGAGCGGTGGCGGTGGGGAGGCCCGAGGCAGTGACACGGGGTAGGGTTGTAGCTCCTCC

        AlaLeuSerThrThrGlyGluIleProPheTyrGlyLysAlaIleProLeuGluValIle
2221  TTGCTCTGTCCACCACCGGAGAGATCCCTTTTTACGGCAAGGCTATCCCCCTCGAAGTAA
      AACGAGACAGGTGGTGGCCTCTCTAGGGAAAAATGCCGTTCCGATAGGGGGAGCTTCATT

        LysGlyGlyArgHisLeuIlePheCysHisSerLysLysLysCysAspGluLeuAlaAla
2281  TCAAGGGGGGGAGACATCTCATCTTCTGTCATTCAAAGAAGAAGTGCGACGAACTCGCCG
      AGTTCCCCCCCCTCTGTAGAGTAGAAGACAGTAAGTTTCTTCTTCACGCTGCTTGAGCGGC

        LysLeuValAlaLeuGlyIleAsnAlaValAlaTyrTyrArgGlyLeuAspValSerVal
2341  CAAAGCTGGTCGCATTGGGCATCAATGCCGTGGCCTACTACCGCGGTCTTGACGTGTCCG
      GTTTCGACCAGCGTAACCCGTAGTTACGGCACCGGATGATGGCGCCAGAACTGCACAGGC

        IleProThrSerGlyAspValValValValAlaThrAspAlaLeuMetThrGlyTyrThr
2401  TCATCCCGACCAGCGGCGATGTTGTCGTCGTGGCAACCGATGCCCTCATGACCGGCTATA
      AGTAGGGCTGGTCGCCGCTACAACAGCAGCACCGTTGGCTACGGGAGTACTGGCCGATAT

        GlyAspPheAspSerValIleAspCysAsnThrCysValThrGlnThrValAspPheSer
2461  CCGGCGACTTCGACTCGGTGATAGACTGCAATACGTGTGTCACCCAGACAGTCGATTTCA
      GGCCGCTGAAGCTGAGCCACTATCTGACGTTATGCACACAGTGGGTCTGTCAGCTAAAGT

        LeuAspProThrPheThrIleGluThrIleThrLeuProGlnAspAlaValSerArgThr
2521  GCCTTGACCCTACCTTCACCATTGAGACAATCACGCTCCCCCAGGATGCTGTCTCCCGCA
      CGGAACTGGGATGGAAGTGGTAACTCTGTTAGTGCGAGGGGGTCCTACGACAGAGGGCGT

        GlnArgArgGlyArgThrGlyArgGlyLysProGlyIleTyrArgPheValAlaProGly
2581  CTCAACGTCGGGGCAGGACTGGCAGGGGGAAGCCAGGCATCTACAGATTTGTGGCACCGG
      GAGTTGCAGCCCCGTCCTGACCGTCCCCCTTCGGTCCGTAGATGTCTAAACACCGTGGCC

        GluArgProSerGlyMetPheAspSerSerValLeuCysGluCysTyrAspAlaGlyCys
2641  GGGAGCGCCCCTCCGGCATGTTCGACTCGTCCGTCCTCTGTGAGTGCTATGACGCAGGCT
      CCCTCGCGGGGAGGCCGTACAAGCTGAGCAGGCAGGAGACACTCACGATACTGCGTCCGA

        AlaTrpTyrGluLeuThrProAlaGluThrThrValArgLeuArgAlaTyrMetAsnThr
2701  GTGCTTGGTATGAGCTCACGCCCGCCGAGACTACAGTTAGGCTACGAGCGTACATGAACA
      CACGAACCATACTCGAGTGCGGGCGGCTCTGATGTCAATCCGATGCTCGCATGTACTTGT

        ProGlyLeuProValCysGlnAspHisLeuGluPheTrpGluGlyValPheThrGlyLeu
2761  CCCCGGGGCTTCCCGTGTGCCAGGACCATCTTGAATTTTGGGAGGGCGTCTTTACAGGCC
      GGGGCCCCGAAGGGCACACGGTCCTGGTAGAACTTAAAACCCTCCCGCAGAAATGTCCGG

        ThrHisIleAspAlaHisPheLeuSerGlnThrLysGlnSerGlyGluAsnLeuProTyr
2821  TCACTCATATAGATGCCCACTTTCTATCCCAGACAAAGCAGAGTGGGGAGAACCTTCCTT
      AGTGAGTATATCTACGGGTGAAAGATAGGGTCTGTTTCGTCTCACCCCTCTTGGAAGGAA

        LeuValAlaTyrGlnAlaThrValCysAlaArgAlaGlnAlaProProProSerTrpAsp
2881  ACCTGGTAGCGTACCAAGCCACCGTGTGCGCTAGGGCTCAAGCCCCTCCCCCATCGTGGG
      TGGACCATCGCATGGTTCGGTGGCACACGCGATCCCGAGTTCGGGGAGGGGGTAGCACCC
```

FIG. 32-3

```
          GlnMetTrpLysCysLeuIleArgLeuLysProThrLeuHisGlyProThrProLeuLeu
2941      ACCAGATGTGGAAGTGTTTGATTCGCCTCAAGCCCACCCTCCATGGGCCAACACCCCTGC
          TGGTCTACACCTTCACAAACTAAGCGGAGTTCGGGTGGGAGGTACCCGGTTGTGGGGACG

          TyrArgLeuGlyAlaValGlnAsnGluIleThrLeuThrHisProValThrLysTyrIle
3001      TATACAGACTGGGCGCTGTTCAGAATGAAATCACCCTGACGCACCCAGTCACCAAATACA
          ATATGTCTGACCCGCGACAAGTCTTACTTTAGTGGGACTGCGTGGGTCAGTGGTTTATGT

          MetThrCysMetSerAlaAspLeuGluValValThrSerThrTrpValLeuValGlyGly
3061      TCATGACATGCATGTCGGCCGACCTGGAGGTCGTCACGAGCACCTGGGTGCTCGTTGGCG
          AGTACTGTACGTACAGCCGGCTGGACCTCCAGCAGTGCTCGTGGACCCACGAGCAACCGC

          ValLeuAlaAlaLeuAlaAlaTyrCysLeuSerThrGlyCysValValIleValGlyArg
3121      GCGTCCTGGCTGCTTTGGCCGCGTATTGCCTGTCAACAGGCTGCGTGGTCATAGTGGGCA
          CGCAGGACCGACGAAACCGGCGCATAACGGACAGTTGTCCGACGCACCAGTATCACCCGT

          ValValLeuSerGlyLysProAlaIleIleProAspArgGluValLeuTyrArgGluPhe
3181      GGGTCGTCTTGTCCGGGAAGCCGGCAATCATACCTGACAGGGAAGTCCTCTACCGAGAGT
          CCCAGCAGAACAGGCCCTTCGGCCGTTAGTATGGACTGTCCCTTCAGGAGATGGCTCTCA

          AspGluMetGluGluCysSerGlnHisLeuProTyrIleGluGlnGlyMetMetLeuAla
3241      TCGATGAGATGGAAGAGTGCTCTCAGCACTTACCGTACATCGAGCAAGGGATGATGCTCG
          AGCTACTCTACCTTCTCACGAGAGTCGTGAATGGCATGTAGCTCGTTCCCTACTACGAGC

          GluGlnPheLysGlnLysAlaLeuGlyLeuLeuGlnThrAlaSerArgGlnAlaGluVal
3301      CCGAGCAGTTCAAGCAGAAGGCCCTCGGCCTCCTGCAGACCGCGTCCCGTCAGGCAGAGG
          GGCTCGTCAAGTTCGTCTTCCGGGAGCCGGAGGACGTCTGGCGCAGGGCAGTCCGTCTCC

          IleAlaProAlaValGlnThrAsnTrpGlnLysLeuGluThrPheTrpAlaLysHisMet
3361      TTATCGCCCCTGCTGTCCAGACCAACTGGCAAAAACTCGAGACCTTCTGGGCGAAGCATA
          AATAGCGGGGACGACAGGTCTGGTTGACCGTTTTTGAGCTCTGGAAGACCCGCTTCGTAT

          TrpAsnPheIleSerGlyIleGlnTyrLeuAlaGlyLeuSerThrLeuProGlyAsnPro
3421      TGTGGAACTTCATCAGTGGGATACAATACTTGGCGGGCTTGTCAACGCTGCCTGGTAACC
          ACACCTTGAAGTAGTCACCCTATGTTATGAACCGCCCGAACAGTTGCGACGGACCATTGG

          AlaIleAlaSerLeuMetAlaPheThrAlaAlaValThrSerProLeuThrThrSerGln
3481      CCGCCATTGCTTCATTGATGGCTTTTACAGCTGCTGTCACCAGCCCACTAACCACTAGCC
          GGCGGTAACGAAGTAACTACCGAAAATGTCGACGACAGTGGTCGGGTGATTGGTGATCGG

          ThrLeuLeuPheAsnIleLeuGlyGlyTrpValAlaAlaGlnLeuAlaAlaProGlyAla
3541      AAACCCTCCTCTTCAACATATTGGGGGGGTGGGTGGCTGCCCAGCTCGCCGCCCCCGGTG
          TTTGGGAGGAGAAGTTGTATAACCCCCCCACCCACCGACGGGTCGAGCGGCGGGGGCCAC

          AlaThrAlaPheValGlyAlaGlyLeuAlaGlyAlaAlaIleGlySerValGlyLeuGly
3601      CCGCTACTGCCTTTGTGGGCGCTGGCTTAGCTGGCGCCGCCATCGGCAGTGTTGGACTGG
          GGCGATGACGGAAACACCCGCGACCGAATCGACCGCGGCGGTAGCCGTCACAACCTGACC

          LysValLeuIleAspIleLeuAlaGlyTyrGlyAlaGlyValAlaGlyAlaLeuValAla
3661      GGAAGGTCCTCATAGACATCCTTGCAGGGTATGGCGCGGGCGTGGCGGGAGCTCTTGTGG
          CCTTCCAGGAGTATCTGTAGGAACGTCCCATACCGCGCCCGCACCGCCCTCGAGAACACC

          PheLysIleMetSerGlyGluValProSerThrGluAspLeuValAsnLeuLeuProAla
3721      CATTCAAGATCATGAGCGGTGAGGTCCCCTCCACGGAGGACCTGGTCAATCTACTGCCCG
          GTAAGTTCTAGTACTCGCCACTCCAGGGGAGGTGCCTCCTGGACCAGTTAGATGACGGGC

          IleLeuSerProGlyAlaLeuValValGlyValValCysAlaAlaIleLeuArgArgHis
3781      CCATCCTCTCGCCCCGGAGCCCTCGTAGTCGGCGTGGTCTGTGCAGCAATACTGCGCCGGC
          GGTAGGAGAGCGGGCCTCGGGAGCATCAGCCGCACCAGACACGTCGTTATGACGCGGCCG

          ValGlyProGlyGluGlyAlaValGlnTrpMetAsnArgLeuIleAlaPheAlaSerArg
3841      ACGTTGGCCCGGGCGAGGGGGCAGTGCAGTGGATGAACCGGCTGATAGCCTTCGCCTCCC
          TGCAACCGGGCCCGCTCCCCCGTCACGTCACCTACTTGGCCGACTATCGGAAGCGGAGGG

          GlyAsnHisValSerProThrHisTyrValProGluSerAspAlaAlaAlaArgValThr
```

FIG. 32-4

```
3901  GGGGGAACCATGTTTCCCCCACGCACTACGTGCCGGAGAGCGATGCAGCTGCCCGCGTCA
      CCCCCTTGGTACAAAGGGGGTGCGTGATGCACGGCCTCTCGCTACGTCGACGGGCGCAGT

        AlaIleLeuSerSerLeuThrValThrGlnLeuLeuArgArgLeuHisGlnTrpIleSer
3961  CTGCCATACTCAGCAGCCTCACTGTAACCCAGCTCCTGAGGCGACTGCACCAGTGGATAA
      GACGGTATGAGTCGTCGGAGTGACATTGGGTCGAGGACTCCGCTGACGTGGTCACCTATT

        SerGluCysThrThrProCysSerGlySerTrpLeuArgAspIleTrpAspTrpIleCys
4021  GCTCGGAGTGTACCACTCCATGCTCCGGTTCCTGGCTAAGGGACATCTGGGACTGGATAT
      CGAGCCTCACATGGTGAGGTACGAGGCCAAGGACCGATTCCCTGTAGACCCTGACCTATA

        GluValLeuSerAspPheLysThrTrpLeuLysAlaLysLeuMetProGlnLeuProGly
4081  GCGAGGTGTTGAGCGACTTTAAGACCTGGCTAAAAGCTAAGCTCATGCCACAGCTGCCTG
      CGCTCCACAACTCGCTGAAATTCTGGACCGATTTTCGATTCGAGTACGGTGTCGACGGAC

        IleProPheValSerCysGlnArgGlyTyrLysGlyValTrpArgValAspGlyIleMet
4141  GGATCCCCTTTGTGTCCTGCCAGCGCGGGTATAAGGGGGTCTGGCGAGTGGACGGCATCA
      CCTAGGGGAAACACAGGACGGTCGCGCCCATATTCCCCCAGACCGCTCACCTGCCGTAGT

        HisThrArgCysHisCysGlyAlaGluIleThrGlyHisValLysAsnGlyThrMetArg
4201  TGCACACTCGCTGCCACTGTGGAGCTGAGATCACTGGACATGTCAAAAACGGGACGATGA
      ACGTGTGAGCGACGGTGACACCTCGACTCTAGTGACCTGTACAGTTTTTGCCCTGCTACT

        IleValGlyProArgThrCysArgAsnMetTrpSerGlyThrPheProIleAsnAlaTyr
4261  GGATCGTCGGTCCTAGGACCTGCAGGAACATGTGGAGTGGGACCTTCCCCATTAATGCCT
      CCTAGCAGCCAGGATCCTGGACGTCCTTGTACACCTCACCCTGGAAGGGGTAATTACGGA

        ThrThrGlyProCysThrProLeuProAlaProAsnTyrThrPheAlaLeuTrpArgVal
4321  ACACCACGGGCCCCTGTACCCCCCTTCCTGCGCCGAACTACACGTTCGCGCTATGGAGGG
      TGTGGTGCCCGGGGACATGGGGGGAAGGACGCGGCTTGATGTGCAAGCGCGATACCTCCC

        SerAlaGluGluTyrValGluIleArgGlnValGlyAspPheHisTyrValThrGlyMet
4381  TGTCTGCAGAGGAATATGTGGAGATAAGGCAGGTGGGGGACTTCCACTACGTGACGGGTA
      ACAGACGTCTCCTTATACACCTCTATTCCGTCCACCCCCTGAAGGTGATGCACTGCCCAT

        ThrThrAspAsnLeuLysCysProCysGlnValProSerProGluPhePheThrGluLeu
4441  TGACTACTGACAATCTCAAATGCCCGTGCCAGGTCCCATCGCCCGAATTTTTCACAGAAT
      ACTGATGACTGTTAGAGTTTACGGGCACGGTCCAGGGTAGCGGGCTTAAAAAGTGTCTTA

        AspGlyValArgLeuHisArgPheAlaProProCysLysProLeuLeuArgGluGluVal
4501  TGGACGGGGTGCGCCTACATAGGTTTGCGCCCCCCTGCAAGCCCTTGCTGCGGGAGGAGG
      ACCTGCCCCACGCGGATGTATCCAAACGCGGGGGGACGTTCGGGAACGACGCCCTCCTCC

        SerPheArgValGlyLeuHisGluTyrProValGlySerGlnLeuProCysGluProGlu
4561  TATCATTCAGAGTAGGACTCCACGAATACCCGGTAGGGTCGCAATTACCTTGCGAGCCCG
      ATAGTAAGTCTCATCCTGAGGTGCTTATGGGCCATCCCAGCGTTAATGGAACGCTCGGGC

        ProAspValAlaValLeuThrSerMetLeuThrAspProSerHisIleThrAlaGluAla
4621  AACCGGACGTGGCCGTGTTGACGTCCATGCTCACTGATCCCTCCCATATAACAGCAGAGG
      TTGGCCTGCACCGGCACAACTGCAGGTACGAGTGACTAGGGAGGGTATATTGTCGTCTCC

        AlaGlyArgArgLeuAlaArgGlySerProProSerValAlaSerSerSerAlaSerGln
4681  CGGCCGGGCGAAGGTTGGCGAGGGGATCACCCCCCTCTGTGGCCAGCTCCTCGGCTAGCC
      GCCGGCCCGCTTCCAACCGCTCCCCTAGTGGGGGGAGACACCGGTCGAGGAGCCGATCGG

        LeuSerAlaProSerLeuLysAlaThrCysThrAlaAsnHisAspSerProAspAlaGlu
4741  AGCTATCCGCTCCATCTCTCAAGGCAACTTGCACCGCTAACCATGACTCCCCTGATGCTG
      TCGATAGGCGAGGTAGAGAGTTCCGTTGAACGTGGCGATTGGTACTGAGGGGACTACGAC

        LeuIleGluAlaAsnLeuLeuTrpArgGlnGluMetGlyGlyAsnIleThrArgValGlu
4801  AGCTCATAGAGGCCAACCTCCTATGGAGGCAGGAGATGGGCGGCAACATCACCAGGGTTG
      TCGAGTATCTCCGGTTGGAGGATACCTCCGTCCTCTACCCGCCGTTGTAGTGGTCCCAAC

        SerGluAsnLysValValIleLeuAspSerPheAspProLeuValAlaGluGluAspGlu
4861  AGTCAGAAAACAAAGTGGTGATTCTGGACTCCTTCGATCCGCTTGTGGCGGAGGAGGACG
      TCAGTCTTTTGTTTCACCACTAAGACCTGAGGAAGCTAGGCGAACACCGCCTCCTCCTGC
```

FIG. 32-5

```
                ArgGluIleSerValProAlaGluIleLeuArgLysSerArgArgPheAlaGlnAlaLeu
4921  AGCGGGAGATCTCCGTACCCGCAGAAATCCTGCGGAAGTCTCGGAGATTCGCCCAGGCCC
      TCGCCCTCTAGAGGCATGGGCGTCTTTAGGACGCCTTCAGAGCCTCTAAGCGGGTCCGGG

                ProValTrpAlaArgProAspTyrAsnProProLeuValGluThrTrpLysLysProAsp
4981  TGCCCGTTTGGGCGCGGCCGGACTATAACCCCCCGCTAGTGGAGACGTGGAAAAAGCCCG
      ACGGGCAAACCCGCGCCGGCCTGATATTGGGGGGCGATCACCTCTGCACCTTTTTCGGGC

                TyrGluProProValValHisGlyCysProLeuProProProLysSerProProValPro
5041  ACTACGAACCACCTGTGGTCCATGGCTGTCCGCTTCCACCTCCAAAGTCCCCTCCTGTGC
      TGATGCTTGGTGGACACCAGGTACCGACAGGCGAAGGTGGAGGTTTCAGGGGAGGACACG

                ProProArgLysLysArgThrValValLeuThrGluSerThrLeuSerThrAlaLeuAla
5101  CTCCGCCTCGGAAGAAGCGGACGGTGGTCCTCACTGAATCAACCCTATCTACTGCCTTGG
      GAGGCGGAGCCTTCTTCGCCTGCCACCAGGAGTGACTTAGTTGGGATAGATGACGGAACC

                GluLeuAlaThrArgSerPheGlySerSerSerThrSerGlyIleThrGlyAspAsnThr
5161  CCGAGCTCGCCACCAGAAGCTTTGGCAGCTCCTCAACTTCCGGCATTACGGGCGACAATA
      GGCTCGAGCGGTGGTCTTCGAAACCGTCGAGGAGTTGAAGGCCGTAATGCCCGCTGTTAT

                ThrThrSerSerGluProAlaProSerGlyCysProProAspSerAspAlaGluSerTyr
5221  CGACAACATCCTCTGAGCCCGCCCCTTCTGGCTGCCCCCCCGACTCCGACGCTGAGTCCT
      GCTGTTGTAGGAGACTCGGGCGGGGAAGACCGACGGGGGGGCTGAGGCTGCGACTCAGGA

                SerSerMetProProLeuGluGlyGluProGlyAspProAspLeuSerAspGlySerTrp
5281  ATTCCTCCATGCCCCCCCTGGAGGGGAGCCTGGGGATCCGGATCTTAGCGACGGGTCAT
      TAAGGAGGTACGGGGGGGACCTCCCCCTCGGACCCCTAGGCCTAGAATCGCTGCCCAGTA

                SerThrValSerSerGluAlaAsnAlaGluAspValValCysCysSerMetSerTyrSer
5341  GGTCAACGGTCAGTAGTGAGGCCAACGCGGAGGATGTCGTGTGCTGCTCAATGTCTTACT
      CCAGTTGCCAGTCATCACTCCGGTTGCGCCTCCTACAGCACACGACGAGTTACAGAATGA

                TrpThrGlyAlaLeuValThrProCysAlaAlaGluGluGlnLysLeuProIleAsnAla
5401  CTTGGACAGGCGCACTCGTCACCCCGTGCGCCGCGGAAGAACAGAAACTGCCCATCAATG
      GAACCTGTCCGCGTGAGCAGTGGGGCACGCGGCGCCTTCTTGTCTTTGACGGGTAGTTAC

                LeuSerAsnSerLeuLeuArgHisHisAsnLeuValTyrSerThrThrSerArgSerAla
5461  CACTAAGCAACTCGTTGCTACGTCACCACAATTTGGTGTATTCCACCACCTCACGCAGTG
      GTGATTCGTTGAGCAACGATGCAGTGGTGTTAAACCACATAAGGTGGTGGAGTGCGTCAC

                CysGlnArgGlnLysLysValThrPheAspArgLeuGlnValLeuAspSerHisTyrGln
5521  CTTGCCAAAGGCAGAAGAAAGTCACATTTGACAGACTGCAAGTTCTGGACAGCCATTACC
      GAACGGTTTCCGTCTTCTTTCAGTGTAAACTGTCTGACGTTCAAGACCTGTCGGTAATGG

                AspValLeuLysGluValLysAlaAlaAlaSerLysValLysAlaAsnLeuLeuSerVal
5581  AGGACGTACTCAAGGAGGTTAAAGCAGCGGCGTCAAAAGTGAAGGCTAACTTGCTATCCG
      TCCTGCATGAGTTCCTCCAATTTCGTCGCCGCAGTTTTCACTTCCGATTGAACGATAGGC

                GluGluAlaCysSerLeuThrProProHisSerAlaLysSerLysPheGlyTyrGlyAla
5641  TAGAGGAAGCTTGCAGCCTGACGCCCCCACACTCAGCCAAATCCAAGTTTGGTTATGGGG
      ATCTCCTTCGAACGTCGGACTGCGGGGGTGTGAGTCGGTTTAGGTTCAAACCAATACCCC

                LysAspValArgCysHisAlaArgLysAlaValThrHisIleAsnSerValTrpLysAsp
5701  CAAAAGACGTCCGTTGCCATGCCAGAAAGGCCGTAACCCACATCAACTCCGTGTGGAAAG
      GTTTTCTGCAGGCAACGGTACGGTCTTTCCGGCATTGGGTGTAGTTGAGGCACACCTTTC

                LeuLeuGluAspAsnValThrProIleAspThrThrIleMetAlaLysAsnGluValPhe
5761  ACCTTCTGGAAGACAATGTAACACCAATAGACACTACCATCATGGCTAAGAACGAGGTTT
      TGGAAGACCTTCTGTTACATTGTGGTTATCTGTGATGGTAGTACCGATTCTTGCTCCAAA

                CysValGlnProGluLysGlyGlyArgLysProAlaArgLeuIleValPheProAspLeu
5821  TCTGCGTTCAGCCTGAGAAGGGGGGTCGTAAGCCAGCTCGTCTCATCGTGTTCCCCGATC
      AGACGCAAGTCGGACTCTTCCCCCCAGCATTCGGTCGAGCAGAGTAGCACAAGGGGCTAG

                GlyValArgValCysGluLysMetAlaLeuTyrAspValValThrLysLeuProLeuAla
```

FIG. 32-6

```
5881  TGGGCGTGCGCGTGTGCGAAAAGATGGCTTTGTACGACGTGGTTACAAAGCTCCCCTTGG
      ACCCGCACGCGCACACGCTTTTCTACCGAAACATGCTGCACCAATGTTTCGAGGGGAACC

        ValMetGlySerSerTyrGlyPheGlnTyrSerProGlyGlnArgValGluPheLeuVal
5941  CCGTGATGGGAAGCTCCTACGGATTCCAATACTCACCAGGACAGCGGGTTGAATTCCTCG
      GGCACTACCCTTCGAGGATGCCTAAGGTTATGAGTGGTCCTGTCGCCCAACTTAAGGAGC

        GlnAlaTrpLysSerLysLysThrProMetGlyPheSerTyrAspThrArgCysPheAsp
6001  TGCAAGCGTGGAAGTCCAAGAAAACCCCAATGGGGTTCTCGTATGATACCCGCTGCTTTG
      ACGTTCGCACCTTCAGGTTCTTTTGGGGTTACCCCAAGAGCATACTATGGGCGACGAAAC

        SerThrValThrGluSerAspIleArgThrGluGluAlaIleTyrGlnCysCysAspLeu
6061  ACTCCACAGTCACTGAGAGCGACATCCGTACGGAGGAGGCAATCTACCAATGTTGTGACC
      TGAGGTGTCAGTGACTCTCGCTGTAGGCATGCCTCCTCCGTTAGATGGTTACAACACTGG

        AspProGlnAlaArgValAlaIleLysSerLeuThrGluArgLeuTyrValGlyGlyPro
6121  TCGACCCCCAAGCCCGCGTGGCCATCAAGTCCCTCACCGAGAGGCTTTATGTTGGGGGCC
      AGCTGGGGGTTCGGGCGCACCGGTAGTTCAGGGAGTGGCTCTCCGAAATACAACCCCCGG

        LeuThrAsnSerArgGlyGluAsnCysGlyTyrArgArgCysArgAlaSerGlyValLeu
6181  CTCTTACCAATTCAAGGGGGGAGAACTGCGGCTATCGCAGGTGCCGCGCGAGCGGCGTAC
      GAGAATGGTTAAGTTCCCCCCTCTTGACGCCGATAGCGTCCACGGCGCGCTCGCCGCATG

        ThrThrSerCysGlyAsnThrLeuThrCysTyrIleLysAlaArgAlaAlaCysArgAla
6241  TGACAACTAGCTGTGGTAACACCCTCACTTGCTACATCAAGGCCCGGGCAGCCTGTCGAG
      ACTGTTGATCGACACCATTGTGGGAGTGAACGATGTAGTTCCGGGCCCGTCGGACAGCTC

        AlaGlyLeuGlnAspCysThrMetLeuValCysGlyAspAspLeuValValIleCysGlu
6301  CCGCAGGGCTCCAGGACTGCACCATGCTCGTGTGTGGCGACGACTTAGTCGTTATCTGTG
      GGCGTCCCGAGGTCCTGACGTGGTACGAGCACACACCGCTGCTGAATCAGCAATAGACAC

        SerAlaGlyValGlnGluAspAlaAlaSerLeuArgAlaPheThrGluAlaMetThrArg
6361  AAAGCGCGGGGGTCCAGGAGGACGCGGCGAGCCTGAGAGCCTTCACGGAGGCTATGACCA
      TTTCGCGCCCCCAGGTCCTCCTGCGCCGCTCGGACTCTCGGAAGTGCCTCCGATACTGGT

        TyrSerAlaProProGlyAspProProGlnProGluTyrAspLeuGluLeuIleThrSer
6421  GGTACTCCGCCCCCCCTGGGGACCCCCCACAACCAGAATACGACTTGGAGCTCATAACAT
      CCATGAGGCGGGGGGGACCCCTGGGGGGTGTTGGTCTTATGCTGAACCTCGAGTATTGTA

        CysSerSerAsnValSerValAlaHisAspGlyAlaGlyLysArgValTyrTyrLeuThr
6481  CATGCTCCTCCAACGTGTCAGTCGCCCACGACGGCGCTGGAAAGAGGGTCTACTACCTCA
      GTACGAGGAGGTTGCACAGTCAGCGGGTGCTGCCGCGACCTTTCTCCCAGATGATGGAGT

        ArgAspProThrThrProLeuAlaArgAlaAlaTrpGluThrAlaArgHisThrProVal
6541  CCCGTGACCCTACAACCCCCCTCGCGAGAGCTGCGTGGGAGACAGCAAGACACACTCCAG
      GGGCACTGGGATGTTGGGGGGAGCGCTCTCGACGCACCCTCTGTCGTTCTGTGTGAGGTC

        AsnSerTrpLeuGlyAsnIleIleMetPheAlaProThrLeuTrpAlaArgMetIleLeu
6601  TCAATTCCTGGCTAGGCAACATAATCATGTTTGCCCCCACACTGTGGGCGAGGATGATAC
      AGTTAAGGACCGATCCGTTGTATTAGTACAAACGGGGGTGTGACACCCGCTCCTACTATG

        MetThrHisPhePheSerValLeuIleAlaArgAspGlnLeuGluGlnAlaLeuAspCys
6661  TGATGACCCATTTCTTTAGCGTCCTTATAGCCAGGGACCAGCTTGAACAGGCCCTCGATT
      ACTACTGGGTAAAGAAATCGCAGGAATATCGGTCCCTGGTCGAACTTGTCCGGGAGCTAA

        GluIleTyrGlyAlaCysTyrSerIleGluProLeuAspLeuProProIleIleGlnArg
6721  GCGAGATCTACGGGGCCTGCTACTCCATAGAACCACTTGATCTACCTCCAATCATTCAAA
      CGCTCTAGATGCCCCGGACGATGAGGTATCTTGGTGAACTAGATGGAGGTTAGTAAGTTT

        Leu
6781  GACTC
      CTGAG
```

FIG. 32-7

FIG. 47-1 COMBINED ORF OF DNAs K9-1 through 15e

```
      GlyCysProGluArgLeuAlaSerCysArgProLeuThrAspPheAspGlnGlyTrpGly
  1   CAGGCTGTCCTGAGAGGCTAGCCAGCTGCCGACCCCTTACCGATTTTGACCAGGGCTGGG
      GTCCGACAGGACTCTCCGATCGGTCGACGGCTGGGGAATGGCTAAAACTGGTCCCGACCC

      ProIleSerTyrAlaAsnGlySerGlyProAspGlnArgProTyrCysTrpHisTyrPro
 61   GCCCTATCAGTTATGCCAACGGAAGCGGCCCCGACCAGCGCCCCTACTGCTGGCACTACC
      CGGGATAGTCAATACGGTTGCCTTCGCCGGGGCTGGTCGCGGGGATGACGACCGTGATGG

      ProLysProCysGlyIleValProAlaLysSerValCysGlyProValTyrCysPheThr
121   CCCCAAAACCTTGCGGTATTGTGCCCGCGAAGAGTGTGTGTGGTCCGGTATATTGCTTCA
      GGGGTTTTGGAACGCCATAACACGGGCGCTTCTCACACACACCAGGCCATATAACGAAGT

      ProSerProValValValGlyThrThrAspArgSerGlyAlaProThrTyrSerTrpGly
181   CTCCCAGCCCCGTGGTGGTGGGAACGACCGACAGGTCGGGCGCGCCCACCTACAGCTGGG
      GAGGGTCGGGGCACCACCACCCTTGCTGGCTGTCCAGCCCGCGCGGGTGGATGTCGACCC

      GluAsnAspThrAspValPheValLeuAsnAsnThrArgProProLeuGlyAsnTrpPhe
241   GTGAAAATGATACGGACGTCTTCGTCCTTAACAATACCAGGCCACCGCTGGGCAATTGGT
      CACTTTTACTATGCCTGCAGAAGCAGGAATTGTTATGGTCCGGTGGCGACCCGTTAACCA

      GlyCysThrTrpMetAsnSerThrGlyPheThrLysValCysGlyAlaProProCysVal
301   TCGGTTGTACCTGGATGAACTCAACTGGATTCACCAAAGTGTGCGGAGCGCCTCCTTGTG
      AGCCAACATGGACCTACTTGAGTTGACCTAAGTGGTTTCACACGCCTCGCGGAGGAACAC

      IleGlyGlyAlaGlyAsnAsnThrLeuHisCysProThrAspCysPheArgLysHisPro
361   TCATCGGAGGGGCGGGCAACAACACCCTGCACTGCCCCACTGATTGCTTCCGCAAGCATC
      AGTAGCCTCCCCGCCCGTTGTTGTGGGACGTGACGGGGTGACTAACGAAGGCGTTCGTAG

      AspAlaThrTyrSerArgCysGlySerGlyProTrpIleThrProArgCysLeuValAsp
421   CGGACGCCACATACTCTCGGTGCGGCTCCGGTCCCTGGATCACACCCAGGTGCCTGGTCG
      GCCTGCGGTGTATGAGAGCCACGCCGAGGCCAGGGACCTAGTGTGGGTCCACGGACCAGC

      TyrProTyrArgLeuTrpHisTyrProCysThrIleAsnTyrThrIlePheLysIleArg
481   ACTACCCGTATAGGCTTTGGCATTATCCTTGTACCATCAACTACACCATATTTAAAATCA
      TGATGGGCATATCCGAAACCGTAATAGGAACATGGTAGTTGATGTGGTATAAATTTTAGT

      MetTyrValGlyGlyValGluHisArgLeuGluAlaAlaCysAsnTrpThrArgGlyGlu
541   GGATGTACGTGGGAGGGGTCGAACACAGGCTGGAAGCTGCCTGCAACTGGACGCGGGGCG
      CCTACATGCACCCTCCCCAGCTTGTGTCCGACCTTCGACGGACGTTGACCTGCGCCCCGC

      ArgCysAspLeuGluAspArgAspArgSerGluLeuSerProLeuLeuLeuThrThrThr
601   AACGTTGCGATCTGGAAGACAGGGACAGGTCCGAGCTCAGCCCGTTACTGCTGACCACTA
      TTGCAACGCTAGACCTTCTGTCCCTGTCCAGGCTCGAGTCGGGCAATGACGACTGGTGAT

      GlnTrpGlnValLeuProCysSerPheThrThrLeuProAlaLeuSerThrGlyLeuIle
661   CACAGTGGCAGGTCCTCCCGTGTTCCTTCACAACCCTACCAGCCTTGTCCACCGGCCTCA
      GTGTCACCGTCCAGGAGGGCACAAGGAAGTGTTGGGATGGTCGGAACAGGTGGCCGGAGT

      HisLeuHisGlnAsnIleValAspValGlnTyrLeuTyrGlyValGlySerSerIleAla
721   TCCACCTCCACCAGAACATTGTGGACGTGCAGTACTTGTACGGGGTGGGGTCAAGCATCG
      AGGTGGAGGTGGTCTTGTAACACCTGCACGTCATGAACATGCCCCACCCCAGTTCGTAGC

      SerTrpAlaIleLysTrpGluTyrValValLeuLeuPheLeuLeuLeuAlaAspAlaArg
781   CGTCCTGGGCCATTAAGTGGGAGTACGTCGTTCTCCTGTTCCTTCTGCTTGCAGACGCGC
      GCAGGACCCGGTAATTCACCCTCATGCAGCAAGAGGACAAGGAAGACGAACGTCTGCGCG

      ValCysSerCysLeuTrpMetMetLeuLeuIleSerGlnAlaGluAlaAlaLeuGluAsn
841   GCGTCTGCTCCTGCTTGTGGATGATGCTACTCATATCCCAAGCGGAGGCGGCTTTGGAGA
      CGCAGACGAGGACGAACACCTACTACGATGAGTATAGGGTTCGCCTCCGCCGAAACCTCT

      LeuValIleLeuAsnAlaAlaSerLeuAlaGlyThrHisGlyLeuValSerPheLeuVal
901   ACCTCGTAATACTTAATGCAGCATCCCTGGCCGGGACGCACGGTCTTGTATCCTTCCTCG
      TGGAGCATTATGAATTACGTCGTAGGGACCGGCCCTGCGTGCCAGAACATAGGAAGGAGC
```

```
                PhePheCysPheAlaTrpTyrLeuLysGlyLysTrpValProGlyAlaValTyrThrPhe
961   TGTTCTTCTGCTTTGCATGGTATTTGAAGGGTAAGTGGGTGCCCGGAGCGGTCTACACCT
      ACAAGAAGACGAAACGTACCATAAACTTCCCATTCACCCACGGGCCTCGCCAGATGTGGA

                TyrGlyMetTrpProLeuLeuLeuLeuLeuAlaLeuProGlnArgAlaTyrAlaLeu
1021  TCTACGGGATGTGGCCTCTCCTCCTGCTCCTGTTGGCGTTGCCCCAGCGGGCGTACGCGC
      AGATGCCCTACACCGGAGAGGAGGACGAGGACAACCGCAACGGGGTCGCCCGCATGCGCG

                AspThrGluValAlaAlaSerCysGlyGlyValValLeuValGlyLeuMetAlaLeuThr
1081  TGGACACGGAGGTGGCCGCGTCGTGTGGCGGTGTTGTTCTCGTCGGGTTGATGGCGCTGA
      ACCTGTGCCTCCACCGGCGCAGCACACCGCCACAACAAGAGCAGCCCAACTACCGCGACT

                LeuSerProTyrTyrLysArgTyrIleSerTrpCysLeuTrpTrpLeuGlnTyrPheLeu
1141  CTCTGTCACCATATTACAAGCGCTATATCAGCTGGTGCTTGTGGTGGCTTCAGTATTTTC
      GAGACAGTGGTATAATGTTCGCGATATAGTCGACCACGAACACCACCGAAGTCATAAAAG

                ThrArgValGluAlaGlnLeuHisValTrpIleProProLeuAsnValArgGlyGlyArg
1201  TGACCAGAGTGGAAGCGCAACTGCACGTGTGGATTCCCCCCCTCAACGTCCGAGGGGGGC
      ACTGGTCTCACCTTCGCGTTGACGTGCACACCTAAGGGGGGGAGTTGCAGGCTCCCCCCG

                AspAlaValIleLeuLeuMetCysAlaValHisProThrLeuValPheAspIleThrLys
1261  GCGACGCCGTCATCTTACTCATGTGTGCTGTACACCCGACTCTGGTATTTGACATCACCA
      CGCTGCGGCAGTAGAATGAGTACACACGACATGTGGGCTGAGACCATAAACTGTAGTGGT

                LeuLeuLeuAlaValPheGlyProLeuTrpIleLeuGlnAlaSerLeuLeuLysValPro
1321  AATTGCTGCTGGCCGTCTTCGGACCCCTTTGGATTCTTCAAGCCAGTTTGCTTAAAGTAC
      TTAACGACGACCGGCAGAAGCCTGGGGAAACCTAAGAAGTTCGGTCAAACGAATTTCATG

                TyrPheValArgValGlnGlyLeuLeuArgPheCysAlaLeuAlaArgLysMetIleGly
1381  CCTACTTTGTGCGCGTCCAAGGCCTTCTCCGGTTCTGCGCGTTAGCGCGGAAGATGATCG
      GGATGAAACACGCGCAGGTTCCGGAAGAGGCCAAGACGCGCAATCGCGCCTTCTACTAGC

                GlyHisTyrValGlnMetValIleIleLysLeuGlyAlaLeuThrGlyThrTyrValTyr
1441  GAGGCCATTACGTGCAAATGGTCATCATTAAGTTAGGGGCGCTTACTGGCACCTATGTTT
      CTCCGGTAATGCACGTTTACCAGTAGTAATTCAATCCCCGCGAATGACCGTGGATACAAA

                AsnHisLeuThrProLeuArgAspTrpAlaHisAsnGlyLeuArgAspLeuAlaValAla
1501  ATAACCATCTCACTCCTCTTCGGGACTGGGCGCACAACGGCTTGCGAGATCTGGCCGTGG
      TATTGGTAGAGTGAGGAGAAGCCCTGACCCGCGTGTTGCCGAACGCTCTAGACCGGCACC

                ValGluProValValPheSerGlnMetGluThrLysLeuIleThrTrpGlyAlaAspThr
1561  CTGTAGAGCCAGTCGTCTTCTCCCAAATGGAGACCAAGCTCATCACGTGGGGGGCAGATA
      GACATCTCGGTCAGCAGAAGAGGGTTTACCTCTGGTTCGAGTAGTGCACCCCCCGTCTAT

                AlaAlaCysGlyAspIleIleAsnGlyLeuProValSerAlaArgArgGlyArgGluIle
1621  CCGCCGCGTGCGGTGACATCATCAACGGCTTGCCTGTTTCCGCCCGCAGGGGCCGGGAGA
      GGCGGCGCACGCCACTGTAGTAGTTGCCGAACGGACAAAGGCGGGCGTCCCCGGCCCTCT

                LeuLeuGlyProAlaAspGlyMetValSerLysGlyTrpArgLeuLeuAlaProIleThr
1681  TACTGCTCGGGCCAGCCGATGGAATGGTCTCCAAGGGGTGGAGGTTGCTGGCGCCCATCA
      ATGACGAGCCCGGTCGGCTACCTTACCAGAGGTTCCCCACCTCCAACGACCGCGGGTAGT

                AlaTyrAlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThrSerLeuThrGlyArg
1741  CGGCGTACGCCCAGCAGACAAGGGGCCTCCTAGGGTGCATAATCACCAGCCTAACTGGCC
      GCCGCATGCGGGTCGTCTGTTCCCCGGAGGATCCCACGTATTAGTGGTCGGATTGACCGG

                AspLysAsnGlnValGluGlyGluValGlnIleValSerThrAlaAlaGlnThrPheLeu
1801  GGGACAAAAACCAAGTGGAGGGTGAGGTCCAGATTGTGTCAACTGCTGCCCAAACCTTCC
      CCCTGTTTTTGGTTCACCTCCCACTCCAGGTCTAACACAGTTGACGACGGGTTTGGAAGG

                AlaThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIle
1861  TGGCAACGTGCATCAATGGGGTGTGCTGGACTGTCTACCACGGGGCCGGAACGAGGACCA
      ACCGTTGCACGTAGTTACCCCACACGACCTGACAGATGGTGCCCCGGCCTTGCTCCTGGT

                AlaSerProLysGlyProValIleGlnMetTyrThrAsnValAspGlnAspLeuValGly
```

FIG. 47-2

```
1921  TCGCGTCACCCAAGGGTCCTGTCATCCAGATGTATACCAATGTAGACCAAGACCTTGTGG
      AGCGCAGTGGGTTCCCAGGACAGTAGGTCTACATATGGTTACATCTGGTTCTGGAACACC

        TrpProAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeu
1981  GCTGGCCCGCTCCGCAAGGTAGCCGCTCATTGACACCCTGCACTTGCGGCTCCTCGGACC
      CGACCGGGCGAGGCGTTCCATCGGCGAGTAACTGTGGGACGTGAACGCCGAGGAGCCTGG

        TyrLeuValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGly
2041  TTTACCTGGTCACGAGGCACGCCGATGTCATTCCCGTGCGCCGGCGGGGTGATAGCAGGG
      AAATGGACCAGTGCTCCGTGCGGCTACAGTAAGGGCACGCGGCCGCCCCACTATCGTCCC

        SerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeu
2101  GCAGCCTGCTGTCGCCCCGGCCCATTTCCTACTTGAAAGGCTCCTCGGGGGGTCCGCTGT
      CGTCGGACGACAGCGGGGCCGGGTAAAGGATGAACTTTCCGAGGAGCCCCCCAGGCGACA

        CysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAla
2161  TGTGCCCCGCGGGGCACGCCGTGGGCATATTTAGGGCCGCGGTGTGCACCCGTGGAGTGG
      ACACGGGGCGCCCCGTGCGGCACCCGTATAAATCCCGGCGCCACACGTGGGCACCTCACC

        LysAlaValAspPheIleProValGluAsnLeuGluThrThrMetArgSerProValPhe
2221  CTAAGGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGACAACCATGAGGTCCCCGGTGT
      GATTCCGCCACCTGAAATAGGGACACCTCTTGGATCTCTGTTGGTACTCCAGGGGCCACA

        ThrAspAsnSerSerProProValValProGlnSerPheGlnValAlaHisLeuHisAla
2281  TCACGGATAACTCCTCTCCACCAGTAGTGCCCCAGAGCTTCCAGGTGGCTCACCTCCATG
      AGTGCCTATTGAGGAGAGGTGGTCATCACGGGGTCTCGAAGGTCCACCGAGTGGAGGTAC

        ProThrGlySerGlyLysSerThrLysValProAlaAlaTyrAlaAlaGlnGlyTyrLys
2341  CTCCCACAGGCAGCGGCAAAAGCACCAAGGTCCCGGCTGCATATGCAGCTCAGGGCTATA
      GAGGGTGTCCGTCGCCGTTTTCGTGGTTCCAGGGCCGACGTATACGTCGAGTCCCGATAT

        ValLeuValLeuAsnProSerValAlaAlaThrLeuGlyPheGlyAlaTyrMetSerLys
2401  AGGTGCTAGTACTCAACCCCTCTGTTGCTGCAACACTGGGCTTTGGTGCTTACATGTCCA
      TCCACGATCATGAGTTGGGGAGACAACGACGTTGTGACCCGAAACCACGAATGTACAGGT

        AlaHisGlyIleAspProAsnIleArgThrGlyValArgThrIleThrThrGlySerPro
2461  AGGCTCATGGGATCGATCCTAACATCAGGACCGGGGTGAGAACAATTACCACTGGCAGCC
      TCCGAGTACCCTAGCTAGGATTGTAGTCCTGGCCCCACTCTTGTTAATGGTGACCGTCGG

        IleThrTyrSerThrTyrGlyLysPheLeuAlaAspGlyGlyCysSerGlyGlyAlaTyr
2521  CCATCACGTACTCCACCTACGGCAAGTTCCTTGCCGACGGCGGGTGCTCGGGGGGCGCTT
      GGTAGTGCATGAGGTGGATGCCGTTCAAGGAACGGCTGCCGCCCACGAGCCCCCCGCGAA

        AspIleIleIleCysAspGluCysHisSerThrAspAlaThrSerIleLeuGlyIleGly
2581  ATGACATAATAATTTGTGACGAGTGCCACTCCACGGATGCCACATCCATCTTGGGCATCG
      TACTGTATTATTAAACACTGCTCACGGTGAGGTGCCTACGGTGTAGGTAGAACCCGTAGC

        ThrValLeuAspGlnAlaGluThrAlaGlyAlaArgLeuValValLeuAlaThrAlaThr
2641  GCACTGTCCTTGACCAAGCAGAGACTGCGGGGGCGAGACTGGTTGTGCTCGCCACCGCCA
      CGTGACAGGAACTGGTTCGTCTCTGACGCCCCCGCTCTGACCAACACGAGCGGTGGCGGT

        ProProGlySerValThrValProHisProAsnIleGluGluValAlaLeuSerThrThr
2701  CCCCTCCGGGCTCCGTCACTGTGCCCCATCCCAACATCGAGGAGGTTGCTCTGTCCACCA
      GGGGAGGCCCGAGGCAGTGACACGGGGTAGGGTTGTAGCTCCTCCAACGAGACAGGTGGT

        GlyGluIleProPheTyrGlyLysAlaIleProLeuGluValIleLysGlyGlyArgHis
2761  CCGGAGAGATCCCTTTTTACGGCAAGGCTATCCCCCTCGAAGTAATCAAGGGGGGGAGAC
      GGCCTCTCTAGGGAAAAATGCCGTTCCGATAGGGGGAGCTTCATTAGTTCCCCCCCTCTG

        LeuIlePheCysHisSerLysLysLysCysAspGluLeuAlaAlaLysLeuValAlaLeu
2821  ATCTCATCTTCTGTCATTCAAAGAAGAAGTGCGACGAACTCGCCGCAAAGCTGGTCGCAT
      TAGAGTAGAAGACAGTAAGTTTCTTCTTCACGCTGCTTGAGCGGCGTTTCGACCAGCGTA

        GlyIleAsnAlaValAlaTyrTyrArgGlyLeuAspValSerValIleProThrSerGly
2881  TGGGCATCAATGCCGTGGCCTACTACCGCGGTCTTGACGTGTCCGTCATCCCGACCAGCG
      ACCCGTAGTTACGGCACCGGATGATGGCGCCAGAACTGCACAGGCAGTAGGGCTGGTCGC
```

FIG. 47-3

```
          AspValValValValAlaThrAspAlaLeuMetThrGlyTyrThrGlyAspPheAspSer
2941      GCGATGTTGTCGTCGTGGCAACCGATGCCCTCATGACCGGCTATACCGGCGACTTCGACT
          CGCTACAACAGCAGCACCGTTGGCTACGGGAGTACTGGCCGATATGGCCGCTGAAGCTGA

          ValIleAspCysAsnThrCysValThrGlnThrValAspPheSerLeuAspProThrPhe
3001      CGGTGATAGACTGCAATACGTGTGTCACCCAGACAGTCGATTTCAGCCTTGACCCTACCT
          GCCACTATCTGACGTTATGCACACAGTGGGTCTGTCAGCTAAAGTCGGAACTGGGATGGA

          ThrIleGluThrIleThrLeuProGlnAspAlaValSerArgThrGlnArgArgGlyArg
3061      TCACCATTGAGACAATCACGCTCCCCCAGGATGCTGTCTCCCGCACTCAACGTCGGGGCA
          AGTGGTAACTCTGTTAGTGCGAGGGGGTCCTACGACAGAGGGCGTGAGTTGCAGCCCCGT

          ThrGlyArgGlyLysProGlyIleTyrArgPheValAlaProGlyGluArgProSerGly
3121      GGACTGGCAGGGGGAAGCCAGGCATCTACAGATTTGTGGCACCGGGGGAGCGCCCCTCCG
          CCTGACCGTCCCCCTTCGGTCCGTAGATGTCTAAACACCGTGGCCCCCTCGCGGGGAGGC

          MetPheAspSerSerValLeuCysGluCysTyrAspAlaGlyCysAlaTrpTyrGluLeu
3181      GCATGTTCGACTCGTCCGTCCTCTGTGAGTGCTATGACGCAGGCTGTGCTTGGTATGAGC
          CGTACAAGCTGAGCAGGCAGGAGACACTCACGATACTGCGTCCGACACGAACCATACTCG

          ThrProAlaGluThrThrValArgLeuArgAlaTyrMetAsnThrProGlyLeuProVal
3241      TCACGCCCGCCGAGACTACAGTTAGGCTACGAGCGTACATGAACACCCCGGGGCTTCCCG
          AGTGCGGGCGGCTCTGATGTCAATCCGATGCTCGCATGTACTTGTGGGGCCCCGAAGGGC

          CysGlnAspHisLeuGluPheTrpGluGlyValPheThrGlyLeuThrHisIleAspAla
3301      TGTGCCAGGACCATCTTGAATTTTGGGAGGGCGTCTTTACAGGCCTCACTCATATAGATG
          ACACGGTCCTGGTAGAACTTAAAACCCTCCCGCAGAAATGTCCGGAGTGAGTATATCTAC

          HisPheLeuSerGlnThrLysGlnSerGlyGluAsnLeuProTyrLeuValAlaTyrGln
3361      CCCACTTTCTATCCCAGACAAAGCAGAGTGGGGAGAACCTTCCTTACCTGGTAGCGTACC
          GGGTGAAAGATAGGGTCTGTTTCGTCTCACCCCTCTTGGAAGGAATGGACCATCGCATGG

          AlaThrValCysAlaArgAlaGlnAlaProProProSerTrpAspGlnMetTrpLysCys
3421      AAGCCACCGTGTGCGCTAGGGCTCAAGCCCCTCCCCCATCGTGGGACCAGATGTGGAAGT
          TTCGGTGGCACACGCGATCCCGAGTTCGGGGAGGGGGTAGCACCCTGGTCTACACCTTCA

          LeuIleArgLeuLysProThrLeuHisGlyProThrProLeuLeuTyrArgLeuGlyAla
3481      GTTTGATTCGCCTCAAGCCCACCCTCCATGGGCCAACACCCCTGCTATACAGACTGGGCG
          CAAACTAAGCGGAGTTCGGGTGGGAGGTACCCGGTTGTGGGGACGATATGTCTGACCCGC

          ValGlnAsnGluIleThrLeuThrHisProValThrLysTyrIleMetThrCysMetSer
3541      CTGTTCAGAATGAAATCACCCTGACGCACCCAGTCACCAAATACATCATGACATGCATGT
          GACAAGTCTTACTTTAGTGGGACTGCGTGGGTCAGTGGTTTATGTAGTACTGTACGTACA

          AlaAspLeuGluValValThrSerThrTrpValLeuValGlyGlyValLeuAlaAlaLeu
3601      CGGCCGACCTGGAGGTCGTCACGAGCACCTGGGTGCTCGTTGGCGGCGTCCTGGCTGCTT
          GCCGGCTGGACCTCCAGCAGTGCTCGTGGACCCACGAGCAACCGCCGCAGGACCGACGAA

          AlaAlaTyrCysLeuSerThrGlyCysValValIleValGlyArgValValLeuSerGly
3661      TGGCCGCGTATTGCCTGTCAACAGGCTGCGTGGTCATAGTGGGCAGGGTCGTCTTGTCCG
          ACCGGCGCATAACGGACAGTTGTCCGACGCACCAGTATCACCCGTCCCAGCAGAACAGGC

          LysProAlaIleIleProAspArgGluValLeuTyrArgGluPheAspGluMetGluGlu
3721      GGAAGCCGGCAATCATACCTGACAGGGAAGTCCTCTACCGAGAGTTCGATGAGATGGAAG
          CCTTCGGCCGTTAGTATGGACTGTCCCTTCAGGAGATGGCTCTCAAGCTACTCTACCTTC

          CysSerGlnHisLeuProTyrIleGluGlnGlyMetMetLeuAlaGluGlnPheLysGln
3781      AGTGCTCTCAGCACTTACCGTACATCGAGCAAGGGATGATGCTCGCCGAGCAGTTCAAGC
          TCACGAGAGTCGTGAATGGCATGTAGCTCGTTCCCTACTACGAGCGGCTCGTCAAGTTCG

          LysAlaLeuGlyLeuLeuGlnThrAlaSerArgGlnAlaGluValIleAlaProAlaVal
3841      AGAAGGCCCTCGGCCTCCTGCAGACCGCGTCCCGTCAGGCAGAGGTTATCGCCCCTGCTG
          TCTTCCGGGAGCCGGAGGACGTCTGGCGCAGGGCAGTCCGTCTCCAATAGCGGGGACGAC

          GlnThrAsnTrpGlnLysLeuGluThrPheTrpAlaLysHisMetTrpAsnPheIleSer
```

FIG. 47-4

```
3901  TCCAGACCAACTGGCAAAAACTCGAGACCTTCTGGGCGAAGCATATGTGGAACTTCATCA
      AGGTCTGGTTGACCGTTTTTGAGCTCTGGAAGACCCGCTTCGTATACACCTTGAAGTAGT

        GlyIleGlnTyrLeuAlaGlyLeuSerThrLeuProGlyAsnProAlaIleAlaSerLeu
3961  GTGGGATACAATACTTGGCGGGCTTGTCAACGCTGCCTGGTAACCCCGCCATTGCTTCAT
      CACCCTATGTTATGAACCGCCCGAACAGTTGCGACGGACCATTGGGGCGGTAACGAAGTA

        MetAlaPheThrAlaAlaValThrSerProLeuThrThrSerGlnThrLeuLeuPheAsn
4021  TGATGGCTTTTACAGCTGCTGTCACCAGCCCACTAACCACTAGCCAAACCCTCCTCTTCA
      ACTACCGAAAATGTCGACGACAGTGGTCGGGTGATTGGTGATCGGTTTGGGAGGAGAAGT

        IleLeuGlyGlyTrpValAlaAlaGlnLeuAlaAlaProGlyAlaAlaThrAlaPheVal
4081  ACATATTGGGGGGGTGGGTGGCTGCCCAGCTCGCCGCCCCCGGTGCCGCTACTGCCTTTG
      TGTATAACCCCCCCACCCACCGACGGGTCGAGCGGCGGGGGCCACGGCGATGACGGAAAC

        GlyAlaGlyLeuAlaGlyAlaAlaIleGlySerValGlyLeuGlyLysValLeuIleAsp
4141  TGGGCGCTGGCTTAGCTGGCGCCGCCATCGGCAGTGTTGGACTGGGGAAGGTCCTCATAG
      ACCCGCGACCGAATCGACCGCGGCGGTAGCCGTCACAACCTGACCCCTTCCAGGAGTATC

        IleLeuAlaGlyTyrGlyAlaGlyValAlaGlyAlaLeuValAlaPheLysIleMetSer
4201  ACATCCTTGCAGGGTATGGCGCGGGCGTGGCGGGAGCTCTTGTGGCATTCAAGATCATGA
      TGTAGGAACGTCCCATACCGCGCCCGCACCGCCCTCGAGAACACCGTAAGTTCTAGTACT

        GlyGluValProSerThrGluAspLeuValAsnLeuLeuProAlaIleLeuSerProGly
4261  GCGGTGAGGTCCCCTCCACGGAGGACCTGGTCAATCTACTGCCCGCCATCCTCTCGCCCG
      CGCCACTCCAGGGGAGGTGCCTCCTGGACCAGTTAGATGACGGGCGGTAGGAGAGCGGGC

        AlaLeuValValGlyValValCysAlaAlaIleLeuArgArgHisValGlyProGlyGlu
4321  GAGCCCTCGTAGTCGGCGTGGTCTGTGCAGCAATACTGCGCCGGCACGTTGGCCCGGGCG
      CTCGGGAGCATCAGCCGCACCAGACACGTCGTTATGACGCGGCCGTGCAACCGGGCCCGC

        GlyAlaValGlnTrpMetAsnArgLeuIleAlaPheAlaSerArgGlyAsnHisValSer
4381  AGGGGGCAGTGCAGTGGATGAACCGGCTGATAGCCTTCGCCTCCCGGGGGAACCATGTTT
      TCCCCCGTCACGTCACCTACTTGGCCGACTATCGGAAGCGGAGGGCCCCCTTGGTACAAA

        ProThrHisTyrValProGluSerAspAlaAlaAlaArgValThrAlaIleLeuSerSer
4441  CCCCCACGCACTACGTGCCGGAGAGCGATGCAGCTGCCCGCGTCACTGCCATACTCAGCA
      GGGGGTGCGTGATGCACGGCCTCTCGCTACGTCGACGGGCGCAGTGACGGTATGAGTCGT

        LeuThrValThrGlnLeuLeuArgArgLeuHisGlnTrpIleSerSerGluCysThrThr
4501  GCCTCACTGTAACCCAGCTCCTGAGGCGACTGCACCAGTGGATAAGCTCGGAGTGTACCA
      CGGAGTGACATTGGGTCGAGGACTCCGCTGACGTGGTCACCTATTCGAGCCTCACATGGT

        ProCysSerGlySerTrpLeuArgAspIleTrpAspTrpIleCysGluValLeuSerAsp
4561  CTCCATGCTCCGGTTCCTGGCTAAGGGACATCTGGGACTGGATATGCGAGGTGTTGAGCG
      GAGGTACGAGGCCAAGGACCGATTCCCTGTAGACCCTGACCTATACGCTCCACAACTCGC

        PheLysThrTrpLeuLysAlaLysLeuMetProGlnLeuProGlyIleProPheValSer
4621  ACTTTAAGACCTGGCTAAAAGCTAAGCTCATGCCACAGCTGCCTGGGATCCCCTTTGTGT
      TGAAATTCTGGACCGATTTTCGATTCGAGTACGGTGTCGACGGACCCTAGGGGAAACACA

        CysGlnArgGlyTyrLysGlyValTrpArgValAspGlyIleMetHisThrArgCysHis
4681  CCTGCCAGCGCGGGTATAAGGGGGTCTGGCGAGTGGACGGCATCATGCACACTCGCTGCC
      GGACGGTCGCGCCCATATTCCCCCAGACCGCTCACCTGCCGTAGTACGTGTGAGCGACGG

        CysGlyAlaGluIleThrGlyHisValLysAsnGlyThrMetArgIleValGlyProArg
4741  ACTGTGGAGCTGAGATCACTGGACATGTCAAAAACGGGACGATGAGGATCGTCGGTCCTA
      TGACACCTCGACTCTAGTGACCTGTACAGTTTTTGCCCTGCTACTCCTAGCAGCCAGGAT

        ThrCysArgAsnMetTrpSerGlyThrPheProIleAsnAlaTyrThrThrGlyProCys
4801  GGACCTGCAGGAACATGTGGAGTGGGACCTTCCCCATTAATGCCTACACCACGGGCCCCT
      CCTGGACGTCCTTGTACACCTCACCCTGGAAGGGGTAATTACGGATGTGGTGCCCGGGGA

        ThrProLeuProAlaProAsnTyrThrPheAlaLeuTrpArgValSerAlaGluGluTyr
4861  GTACCCCCCTTCCTGCGCCGAACTACACGTTCGCGCTATGGAGGGTGTCTGCAGAGGAAT
      CATGGGGGGAAGGACGCGGCTTGATGTGCAAGCGCGATACCTCCCACAGACGTCTCCTTA
```

FIG. 47-5

```
          ValGluIleArgGlnValGlyAspPheHisTyrValThrGlyMetThrThrAspAsnLeu
4921      ATGTGGAGATAAGGCAGGTGGGGGACTTCCACTACGTGACGGGTATGACTACTGACAATC
          TACACCTCTATTCCGTCCACCCCCTGAAGGTGATGCACTGCCCATACTGATGACTGTTAG

          LysCysProCysGlnValProSerProGluPhePheThrGluLeuAspGlyValArgLeu
4981      TCAAATGCCCGTGCCAGGTCCCATCGCCCGAATTTTTCACAGAATTGGACGGGGTGCGCC
          AGTTTACGGGCACGGTCCAGGGTAGCGGGCTTAAAAAGTGTCTTAACCTGCCCCACGCGG

          HisArgPheAlaProProCysLysProLeuLeuArgGluGluValSerPheArgValGly
5041      TACATAGGTTTGCGCCCCCCTGCAAGCCCTTGCTGCGGGAGGAGGTATCATTCAGAGTAG
          ATGTATCCAAACGCGGGGGGACGTTCGGGAACGACGCCCTCCTCCATAGTAAGTCTCATC

          LeuHisGluTyrProValGlySerGlnLeuProCysGluProGluProAspValAlaVal
5101      GACTCCACGAATACCCGGTAGGGTCGCAATTACCTTGCGAGCCCGAACCGGACGTGGCCG
          CTGAGGTGCTTATGGGCCATCCCAGCGTTAATGGAACGCTCGGGCTTGGCCTGCACCGGC

          LeuThrSerMetLeuThrAspProSerHisIleThrAlaGluAlaAlaGlyArgArgLeu
5161      TGTTGACGTCCATGCTCACTGATCCCTCCCATATAACAGCAGAGGCGGCCGGGCGAAGGT
          ACAACTGCAGGTACGAGTGACTAGGGAGGGTATATTGTCGTCTCCGCCGGCCCGCTTCCA

          AlaArgGlySerProProSerValAlaSerSerSerAlaSerGlnLeuSerAlaProSer
5221      TGGCGAGGGGATCACCCCCCTCTGTGGCCAGCTCCTCGGCTAGCCAGCTATCCGCTCCAT
          ACCGCTCCCCTAGTGGGGGGAGACACCGGTCGAGGAGCCGATCGGTCGATAGGCGAGGTA

          LeuLysAlaThrCysThrAlaAsnHisAspSerProAspAlaGluLeuIleGluAlaAsn
5281      CTCTCAAGGCAACTTGCACCGCTAACCATGACTCCCCTGATGCTGAGCTCATAGAGGCCA
          GAGAGTTCCGTTGAACGTGGCGATTGGTACTGAGGGGACTACGACTCGAGTATCTCCGGT

          LeuLeuTrpArgGlnGluMetGlyGlyAsnIleThrArgValGluSerGluAsnLysVal
5341      ACCTCCTATGGAGGCAGGAGATGGGCGGCAACATCACCAGGGTTGAGTCAGAAAACAAAG
          TGGAGGATACCTCCGTCCTCTACCCGCCGTTGTAGTGGTCCCAACTCAGTCTTTTGTTTC

          ValIleLeuAspSerPheAspProLeuValAlaGluGluAspGluArgGluIleSerVal
5401      TGGTGATTCTGGACTCCTTCGATCCGCTTGTGGCGGAGGAGGACGAGCGGGAGATCTCCG
          ACCACTAAGACCTGAGGAAGCTAGGCGAACACCGCCTCCTCCTGCTCGCCCTCTAGAGGC

          ProAlaGluIleLeuArgLysSerArgArgPheAlaGlnAlaLeuProValTrpAlaArg
5461      TACCCGCAGAAATCCTGCGGAAGTCTCGGAGATTCGCCCAGGCCCTGCCCGTTTGGGCGC
          ATGGGCGTCTTTAGGACGCCTTCAGAGCCTCTAAGCGGGTCCGGGACGGGCAAACCCGCG

          ProAspTyrAsnProProLeuValGluThrTrpLysLysProAspTyrGluProProVal
5521      GGCCGGACTATAACCCCCCGCTAGTGGAGACGTGGAAAAAGCCCGACTACGAACCACCTG
          CCGGCCTGATATTGGGGGGCGATCACCTCTGCACCTTTTTCGGGCTGATGCTTGGTGGAC

          ValHisGlyCysProLeuProProProLysSerProProValProProProArgLysLys
5581      TGGTCCATGGCTGTCCGCTTCCACCTCCAAAGTCCCCTCCTGTGCCTCCGCCTCGGAAGA
          ACCAGGTACCGACAGGCGAAGGTGGAGGTTTCAGGGGAGGACACGGAGGCGGAGCCTTCT

          ArgThrValValLeuThrGluSerThrLeuSerThrAlaLeuAlaGluLeuAlaThrArg
5641      AGCGGACGGTGGTCCTCACTGAATCAACCCTATCTACTGCCTTGGCCGAGCTCGCCACCA
          TCGCCTGCCACCAGGAGTGACTTAGTTGGGATAGATGACGGAACCGGCTCGAGCGGTGGT

          SerPheGlySerSerSerThrSerGlyIleThrGlyAspAsnThrThrThrSerSerGlu
5701      GAAGCTTTGGCAGCTCCTCAACTTCCGGCATTACGGGCGACAATACGACAACATCCTCTG
          CTTCGAAACCGTCGAGGAGTTGAAGGCCGTAATGCCCGCTGTTATGCTGTTGTAGGAGAC

          ProAlaProSerGlyCysProProAspSerAspAlaGluSerTyrSerSerMetProPro
5761      AGCCCGCCCCTTCTGGCTGCCCCCCCGACTCCGACGCTGAGTCCTATTCCTCCATGCCCC
          TCGGGCGGGGAAGACCGACGGGGGGGCTGAGGCTGCGACTCAGGATAAGGAGGTACGGGG

          LeuGluGlyGluProGlyAspProAspLeuSerAspGlySerTrpSerThrValSerSer
5821      CCCTGGAGGGGGAGCCTGGGGATCCGGATCTTAGCGACGGGTCATGGTCAACGGTCAGTA
          GGGACCTCCCCCTCGGACCCCTAGGCCTAGAATCGCTGCCCAGTACCAGTTGCCAGTCAT

          GluAlaAsnAlaGluAspValValCysCysSerMetSerTyrSerTrpThrGlyAlaLeu
```

FIG. 47-6

```
5881  GTGAGGCCAACGCGGAGGATGTCGTGTGCTGCTCAATGTCTTACTCTTGGACAGGCGCAC
      CACTCCGGTTGCGCCTCCTACAGCACACGACGAGTTACAGAATGAGAACCTGTCCGCGTG

        ValThrProCysAlaAlaGluGluGlnLysLeuProIleAsnAlaLeuSerAsnSerLeu
5941  TCGTCACCCCGTGCGCCGCGGAAGAACAGAAACTGCCCATCAATGCACTAAGCAACTCGT
      AGCAGTGGGGCACGCGGCGCCTTCTTGTCTTTGACGGGTAGTTACGTGATTCGTTGAGCA

        LeuArgHisHisAsnLeuValTyrSerThrThrSerArgSerAlaCysGlnArgGlnLys
6001  TGCTACGTCACCACAATTTGGTGTATTCCACCACCTCACGCAGTGCTTGCCAAAGGCAGA
      ACGATGCAGTGGTGTTAAACCACATAAGGTGGTGGAGTGCGTCACGAACGGTTTCCGTCT

        LysValThrPheAspArgLeuGlnValLeuAspSerHisTyrGlnAspValLeuLysGlu
6061  AGAAAGTCACATTTGACAGACTGCAAGTTCTGGACAGCCATTACCAGGACGTACTCAAGG
      TCTTTCAGTGTAAACTGTCTGACGTTCAAGACCTGTCGGTAATGGTCCTGCATGAGTTCC

        ValLysAlaAlaAlaSerLysValLysAlaAsnLeuLeuSerValGluGluAlaCysSer
6121  AGGTTAAAGCAGCGGCGTCAAAAGTGAAGGCTAACTTGCTATCCGTAGAGGAAGCTTGCA
      TCCAATTTCGTCGCCGCAGTTTTCACTTCCGATTGAACGATAGGCATCTCCTTCGAACGT

        LeuThrProProHisSerAlaLysSerLysPheGlyTyrGlyAlaLysAspValArgCys
6181  GCCTGACGCCCCCACACTCAGCCAAATCCAAGTTTGGTTATGGGGCAAAAGACGTCCGTT
      CGGACTGCGGGGGTGTGAGTCGGTTTAGGTTCAAACCAATACCCCGTTTTCTGCAGGCAA

        HisAlaArgLysAlaValThrHisIleAsnSerValTrpLysAspLeuLeuGluAspAsn
6241  GCCATGCCAGAAAGGCCGTAACCCACATCAACTCCGTGTGGAAAGACCTTCTGGAAGACA
      CGGTACGGTCTTTCCGGCATTGGGTGTAGTTGAGGCACACCTTTCTGGAAGACCTTCTGT

        ValThrProIleAspThrThrIleMetAlaLysAsnGluValPheCysValGlnProGlu
6301  ATGTAACACCAATAGACACTACCATCATGGCTAAGAACGAGGTTTTCTGCGTTCAGCCTG
      TACATTGTGGTTATCTGTGATGGTAGTACCGATTCTTGCTCCAAAAGACGCAAGTCGGAC

        LysGlyGlyArgLysProAlaArgLeuIleValPheProAspLeuGlyValArgValCys
6361  AGAAGGGGGGTCGTAAGCCAGCTCGTCTCATCGTGTTCCCCGATCTGGGCGTGCGCGTGT
      TCTTCCCCCCAGCATTCGGTCGAGCAGAGTAGCACAAGGGGCTAGACCCGCACGCGCACA

        GluLysMetAlaLeuTyrAspValValThrLysLeuProLeuAlaValMetGlySerSer
6421  GCGAAAAGATGGCTTTGTACGACGTGGTTACAAAGCTCCCCTTGGCCGTGATGGGAAGCT
      CGCTTTTCTACCGAAACATGCTGCACCAATGTTTCGAGGGGAACCGGCACTACCCTTCGA

        TyrGlyPheGlnTyrSerProGlyGlnArgValGluPheLeuValGlnAlaTrpLysSer
6481  CCTACGGATTCCAATACTCACCAGGACAGCGGGTTGAATTCCTCGTGCAAGCGTGGAAGT
      GGATGCCTAAGGTTATGAGTGGTCCTGTCGCCCAACTTAAGGAGCACGTTCGCACCTTCA

        LysLysThrProMetGlyPheSerTyrAspThrArgCysPheAspSerThrValThrGlu
6541  CCAAGAAAACCCCAATGGGGTTCTCGTATGATACCCGCTGCTTTGACTCCACAGTCACTG
      GGTTCTTTTGGGGTTACCCCAAGAGCATACTATGGGCGACGAAACTGAGGTGTCAGTGAC

        SerAspIleArgThrGluGluAlaIleTyrGlnCysCysAspLeuAspProGlnAlaArg
6601  AGAGCGACATCCGTACGGAGGAGGCAATCTACCAATGTTGTGACCTCGACCCCCAAGCCC
      TCTCGCTGTAGGCATGCCTCCTCCGTTAGATGGTTACAACACTGGAGCTGGGGGTTCGGG

        ValAlaIleLysSerLeuThrGluArgLeuTyrValGlyGlyProLeuThrAsnSerArg
6661  GCGTGGCCATCAAGTCCCTCACCGAGAGGCTTTATGTTGGGGGCCCTCTTACCAATTCAA
      CGCACCGGTAGTTCAGGGAGTGGCTCTCCGAAATACAACCCCCGGGAGAATGGTTAAGTT

        GlyGluAsnCysGlyTyrArgArgCysArgAlaSerGlyValLeuThrThrSerCysGly
6721  GGGGGGAGAACTGCGGCTATCGCAGGTGCCGCGCGAGCGGCGTACTGACAACTAGCTGTG
      CCCCCCTCTTGACGCCGATAGCGTCCACGGCGCGCTCGCCGCATGACTGTTGATCGACAC

        AsnThrLeuThrCysTyrIleLysAlaArgAlaAlaCysArgAlaAlaGlyLeuGlnAsp
6781  GTAACACCCTCACTTGCTACATCAAGGCCCGGGCAGCCTGTCGAGCCGCAGGGCTCCAGG
      CATTGTGGGAGTGAACGATGTAGTTCCGGGCCCGTCGGACAGCTCGGCGTCCCGAGGTCC

        CysThrMetLeuValCysGlyAspAspLeuValValIleCysGluSerAlaGlyValGln
6841  ACTGCACCATGCTCGTGTGTGGCGACGACTTAGTCGTTATCTGTGAAAGCGCGGGGGTCC
      TGACGTGGTACGAGCACACACCGCTGCTGAATCAGCAATAGACACTTTCGCGCCCCCAGG
```

FIG. 47-7

```
          GluAspAlaAlaSerLeuArgAlaPheThrGluAlaMetThrArgTyrSerAlaProPro
6901      AGGAGGACGCGGCGAGCCTGAGAGCCTTCACGGAGGCTATGACCAGGTACTCCGCCCCCC
          TCCTCCTGCGCCGCTCGGACTCTCGGAAGTGCCTCCGATACTGGTCCATGAGGCGGGGGG

          GlyAspProProGlnProGluTyrAspLeuGluLeuIleThrSerCysSerSerAsnVal
6961      CTGGGGACCCCCCACAACCAGAATACGACTTGGAGCTCATAACATCATGCTCCTCCAACG
          GACCCCTGGGGGGTGTTGGTCTTATGCTGAACCTCGAGTATTGTAGTACGAGGAGGTTGC

          SerValAlaHisAspGlyAlaGlyLysArgValTyrTyrLeuThrArgAspProThrThr
7021      TGTCAGTCGCCCACGACGGCGCTGGAAAGAGGGTCTACTACCTCACCCGTGACCCTACAA
          ACAGTCAGCGGGTGCTGCCGCGACCTTTCTCCCAGATGATGGAGTGGGCACTGGGATGTT

          ProLeuAlaArgAlaAlaTrpGluThrAlaArgHisThrProValAsnSerTrpLeuGly
7081      CCCCCCTCGCGAGAGCTGCGTGGGAGACAGCAAGACACACTCCAGTCAATTCCTGGCTAG
          GGGGGGAGCGCTCTCGACGCACCCTCTGTCGTTCTGTGTGAGGTCAGTTAAGGACCGATC

          AsnIleIleMetPheAlaProThrLeuTrpAlaArgMetIleLeuMetThrHisPhePhe
7141      GCAACATAATCATGTTTGCCCCCACACTGTGGGCGAGGATGATACTGATGACCCATTTCT
          CGTTGTATTAGTACAAACGGGGGTGTGACACCCGCTCCTACTATGACTACTGGGTAAAGA

          SerValLeuIleAlaArgAspGlnLeuGluGlnAlaLeuAspCysGluIleTyrGlyAla
7201      TTAGCGTCCTTATAGCCAGGGACCAGCTTGAACAGGCCCTCGATTGCGAGATCTACGGGG
          AATCGCAGGAATATCGGTCCCTGGTCGAACTTGTCCGGGAGCTAACGCTCTAGATGCCCC

          CysTyrSerIleGluProLeuAspLeuProProIleIleGlnArgLeu
7261      CCTGCTACTCCATAGAACCACTTGATCTACCTCCAATCATTCAAAGACTC
          GGACGATGAGGTATCTTGGTGAACTAGATGGAGGTTAGTAAGTTTCTGAG
```

FIG. 47- 8